Author:
Booth, Michael P. S

Title:
Structural studies of the human glyoxylate reductase/hydroxypyruvate reductase (GRHPR) and other proteins

Structural Studies of the Human Glyoxylate Reductase / Hydroxypyruvate Reductase (GRHPR) and other Proteins

Michael P. S. Booth

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Medical and Veterinary Sciences.

Department of Biochemistry, September 2008

Word Count: 56,630

Abstract

Human glyoxylate reductase/hydroxypyruvate reductase (GRHPR) is a D-2-hydroxyacid dehydrogenase found predominantly within the liver, and is essential for the removal of the reactive metabolic product glyoxylate. Primary hyperoxaluria type 2 (PH2) disease, which is characterised by increased urinary oxalate and L-glycerate levels leading to calcium oxalate deposition and decreased renal function, is caused by mutations in the gene for GRHPR.

In this study the first crystal structures of the human GRHPR enzyme are reported. Structures of apo, binary and ternary forms of GRHPR are reported, the latter being the first structure of a true ternary complex of an enzyme from the D-2-hydroxyacid dehydrogenase family. These confirm one of two previously proposed models for substrate binding. The crystal structures have also been used to explain the unusual substrate specificity of the GRHPR family of enzymes and the reduction in GRHPR activity associated with missense mutations that lead to PH2.

Additionally, an enzymatic investigation is also reported and shows that within the cell there is likely to be a fine balance between GRHPR and L-lactate dehydrogenase in their turnover of glyoxylate and hydroxypyruvate substrates. The data suggest that turnover by GRHPR is favoured due to cellular conditions rather then a kinetic preference.

Separately, human sorting nexin 1 (SNX1) - an important component of the endosomal sorting pathway - has also been recombinantly expressed and purified. This includes truncated forms comprising its component phox-homology (PX) and Bin/Amphiphysis/Rvs (BAR) domains. Attempts to produce crystals of these proteins have so far been unsuccessful.

Finally, studies regarding the interaction of plant RKIP/PEBP proteins with their potential binding partners were also initiated. While unsuccessful, this work was useful as the starting point for future studies of these proteins in higher order complexes.

Acknowledgements

Firstly, I must thank my supervisor, Leo Brady, for the opportunity to do my PhD in the UK, even though it has taken a bit longer then originally planned. His continuous assistance and advice throughout my PhD has been much appreciated.

All the members of the lab past and present; Becky, Nick, Tim, Aew, Vicky, Chris, Liz, May and Patrick, thank you for everything, the 4am chats at the synchrotron after none of my crystals have diffracted, the not noticing my singing along to Radio1 in the lab and generally making the lab and office a good place to come to. Particular thanks must go to Becky, whose knowledge of all things crystallographic was matched by her willingness to share it and Nick, who has valiantly stepped into the role over the past year.

Thanks also have to go to my friends, new and old, who have helped with doses of sanity when required and, equally important, the insanity sometimes necessary. Thanks for making me laugh like there is no tomorrow and dance like no one is watching.

Lastly, my family, who wanted me to stay close but knew I had to go. While they are about as far away as they can be they always believed I could do it, even if they were not totally sure what 'it' was. I know I always have their support.

Authors' Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of this thesis has been submitted for any other academic award. Any views expressed in this thesis are those of the author.

SIGNED: ..... DATE: .....2008...

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Å	Angstroms
A_{280nm}	Absorbance at 280 nanometers
Ampl	Amplitude
bp	Base pairs
Cα	Alpha carbon
CAP	Calf Alkaline Phosphatase
CCP4	Collaborative Computational Project No. 4
CCP4mg	CCP4 molecular graphics
CV	Column volumes
Da	Daltons
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
di-C4-PtdIns(3)P	di-butanoyl-phosphatidylinositidol 3-phosphate
di-C4-PtdIns(3,5)P_2	di-butanoyl-phosphatidylinositidol 3,5-bisphosphate
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
FT	Flowering locus T
GSH	Reduced glutathione
GSSH	Oxidised glutathione
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Poly-Histidine tag
LB	Luria Bertani media
IPTG	Isopropyl-β-D-thiogalactopyranoside
LLG	Log Likelihood Gain
MES	2-(*N*-morpholino)ethanesulfonic acid
MWCO	Molecular weight cut-off
NADPH	Nicotinamide adenine dinucleotide phosphate
NADH	Nicotinamide adenine dinucleotide
NCS	Non crystallographic symmetry

NMR	Nuclear magnetic resonance
OD_{600}	Optical density measured at 600 nanometers
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEBP	Phosphatidylethanolamine-binding protein
PEG	Polyethylene glycol
PH	Primary Hyperoxaluria
P_i	Inorganic phosphate (PO_4)
PMSF	Phenylmethylsulphonyl fluoride
PtdInsP	Phosphatidylinositol phosphates
PX beamline	Protein crystallography beamline
RKIP	Raf kinase inhibitor protein
rmsd	Relative mean square deviation
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SNX	Sorting Nexin
TAE	Tris-HCl-acetate-EDTA
TFL1	Terminal flower 1
TLS	Translation/Libration/Screw
TLSMD	TLS Motion Determination
Tris-HCl	Tris-HClhydroxymethylaminomethane
UV	Ultra violet
w/v	Weight for volume
v/v	Volume for volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ε	Extinction coefficient
λ	Wavelength

Amino acids are abbreviated in the standard three or one letter code.

Chapter 1 Introduction

1.1 Overview

The human metabolic enzyme glyoxylate reductase / hydroxypyruvate reductase (GRHPR) is predominantly found in the liver and plays an important role in the metabolism of glyoxylate. Mutations within the gene coding for this enzyme have been found to be the cause of the rare, recessive disorder Primary Hyperoxaluria type 2 (PH2) which is typified by the overproduction of oxalate and L-glycerate (reviewed in Danpure and Purdue, 1995; Danpure and Rumsby, 2004). However, despite its importance in the pathology of PH2, little detailed structural or enzymatic information has been available for GRHPR including the mechanism by which its substrate specificity is achieved. Consequently, studies to investigate the structure and kinetics of this important enzyme have been undertaken and are described in this thesis. Some of these are also presented in the publications arising from this work which are listed in Appendix 1.

Two additional exploratory structural projects have also been undertaken, one involving the human sorting nexin 1 protein involved in endosomal sorting and the other, plant members of the RKIP/PEBP family of proteins. These proteins interact with diverse ligands and the structural determinants of their ligand specificities are unclear. All aspects of these projects, including a background introduction and relevant methodology are included in the appropriate chapters.

1.2 Glyoxylate metabolism

Glyoxylate is a reactive metabolite whose abnormal metabolism results in the over production of oxalate and consequently leads to the pathology of the primary hyperoxalurias (Figure 1.1, Frederick *et al.*, 1963). The source of glyoxylate in human liver has long been unclear, but recent work suggests the majority is derived from the metabolism of hydroxyproline (a component amino acid of collagen) in hepatocyte mitochondria (Knight *et al.*, 2006b).

The key enzymes involved in the further metabolism glyoxylate have been identified as glyoxylate reductase/hydroxypyruvate reductase (GRHPR, Cramer *et al.*, 1999), glycolate oxidase (GO, Richardson and Tolbert, 1961) and alanine:glyoxylate aminotransferase (AGT, Danpure and Jennings, 1986), largely in view of the importance of this pathway in the primary hyperoxalurias.

GRHPR activity is found predominantly within the cytosol of liver cells where it catalyses the reduction of both glyoxylate and hydroxypyruvate using NADPH as a coenzyme. A small component is also found in the mitochondria (Mistry, Danpure and Chalmers, 1988; Cramer *et al.*, 1999). Both AGT (Danpure and Jennings, 1986) and GO (Jones, Morrell and Gould, 2000) are liver specific, peroxisomal enzymes (Figure 1.1). Like GRHPR, a small amount of AGT activity is present in liver mitochondria, but this enzyme (AGT2) has no homology with the peroxisomal form (Baker *et al.*, 2004). The importance of GRHPR and AGT2 activity in the mitochondria is unknown.

Consequently, within a healthy individual glyoxylate is reduced in the cytosol (or mitochondria) to glycolate by GRHPR (Figure 1.1). The glycolate then moves into the peroxisome where it is oxidised back to glyoxylate by GO and a transamination reaction takes place to produce pyruvate and glycine (Figure 1.1). The resulting pyruvate can be recycled back to alanine via a transamination reaction with serine, also catalysed by AGT (Xue *et al.*, 1999). This reaction produces hydroxypyruvate which is in turn reduced to D-glycerate by GRHPR, which subsequently enters the gluconeogenesis pathway. The activity of AGT in the production of hydroxypyruvate has been shown to play a significant role in gluconeogenesis from serine in human liver (Xue *et al.*, 1999).

Why the conversion of glyoxylate to glycolate prior to its entry to the peroxisome is required is unclear, but is likely due to the permeability of this organelle. The permeability of the peroxisomal membrane to glycolate is poorly understood. While rat hepatocyte peroxisomes appear freely permeable to glycolate (Verleur and Wanders, 1993) human hepatocyte peroxisomes are seen to show a small increase in glycolate metabolism when the membrane is disrupted, indicating some transport restriction (Baker *et al.*, 2004). The permeability of the organelle to glyoxylate has not been determined.

The cytoplasmic enzyme L-Lactate Dehydrogenase (L-LDH) can catalyse the reduction of hydroxypyruvate to L-glycerate (Meister, 1952) and the oxidation of glyoxylate to oxalate (Sawaki, Hattori and Yamada, 1967, Figure 1.1). However the role of L-LDH in normal glyoxylate metabolism, when both L-LDH and GRHPR are present, is not clear. This question is partially addressed within this thesis. Although a small amount of oxalate is normally produced in a healthy individual it is unclear if this results either from the activity of L-LDH or GO (Danpure and Purdue, 1995). The ability of L-LDH to catalyse the oxidation of glyoxylate has led to it being proposed as the source of the excess oxalate that results in the pathology associated with a small number of inherited metabolic disorders, the primary hyperoxalurias (Williams and Smith, 1968b).

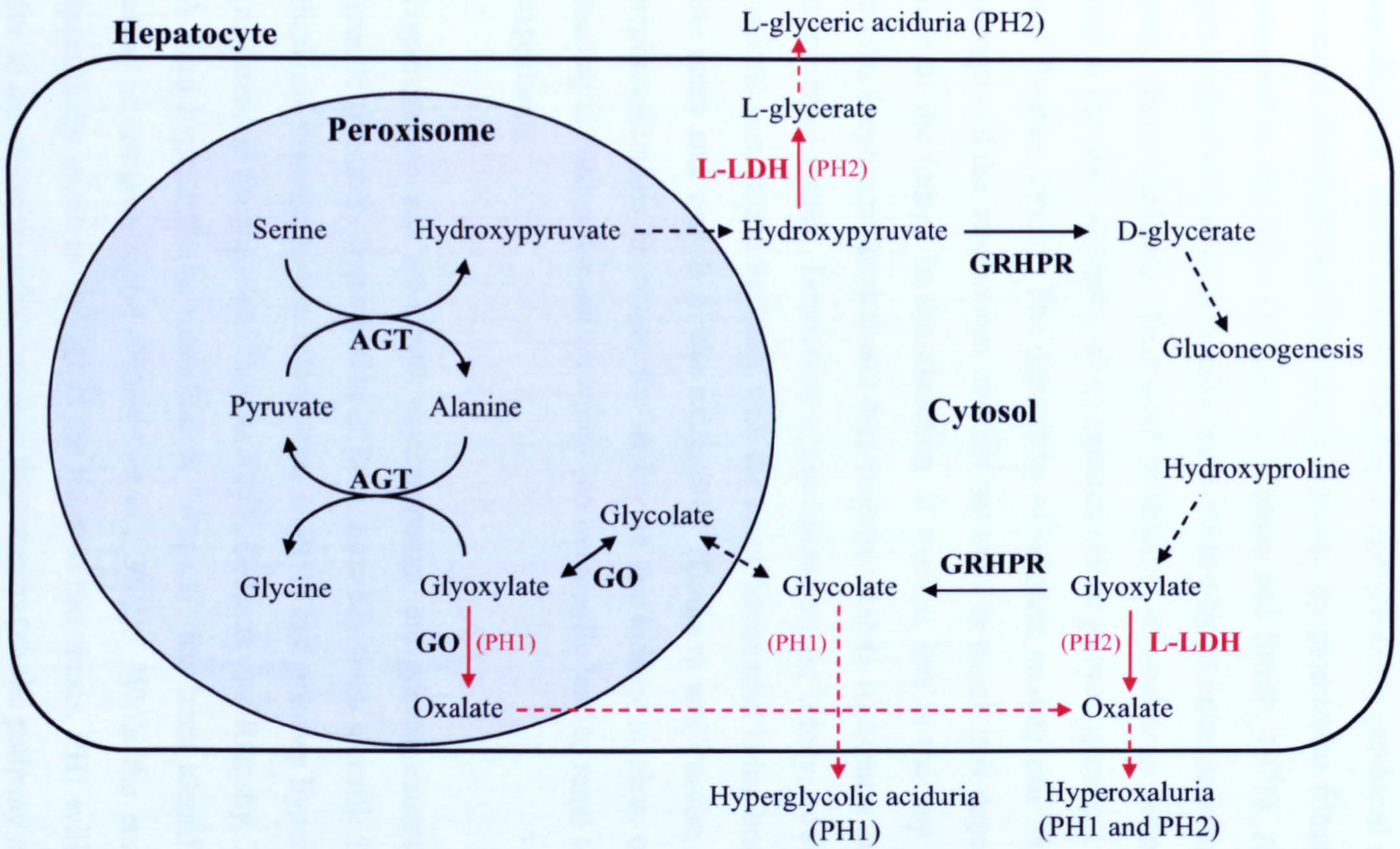


Figure 1.1 Pathways of glyoxylate metabolism in the mammalian liver.

Proposed metabolic paths that dominate during the primary hyperoxalurias are indicated in red. GR: glyoxylate reductase, HPR: hydroxypyruvate reductase, GO: glycolate oxidase, AGT: alanine:glyoxylate aminotransferase, PH1: primary hyperoxaluria 1, PH2: primary hyperoxaluria 2, L-LDH: L-lactate dehydrogenase. Non-direct links or unclear pathways are indicated by dashed lines. Potential metabolism within the mitochondria has been excluded for clarity. Adapted from Danpure and Rumsby, (2004).

1.3 The primary hyperoxalurias

Primary hyperoxalurias are a group of inherited metabolic defects where the predominant feature is the overproduction of oxalic acid (oxalate) which cannot be metabolised. Under normal conditions endogenously produced oxalate (and the small amount absorbed from the diet) is removed by glomerular filtration in the kidney and excreted in the urine (Williams, Johnson and Smith, 1971). An increase in oxalate production, or intake, is usually accompanied by an increase in the level excreted in the urine (hyperoxaluria). Excessive production of oxalate is potentially life-threatening due to the poor solubility of its calcium salt at physiological pH (reviewed in Danpure and Purdue, 1995). The deposition of calcium oxalate can block tubular structures, disrupt cellular interactions and kill the cells in which it is deposited. Because of the role of the kidney in the excretion of oxalate, this is the key site where deposition occurs (nephrocalcinosis) and thus calcium oxalate constitutes the major component of many renal stones. Deposition is not limited to the kidney, however, and can occur at various sites within the body with the most detrimental being bone, the myocardium of the heart and vessels of the extremities (Danpure and Purdue, 1995). Additionally, nephrocalcinosis decreases the ability of the kidney to clear oxalate from the body leading to further deposition which can eventually lead to renal failure and reduced life expectancy.

Hyperoxaluria can have both environmental and genetic causes. A small number of genetic (primary) hyperoxalurias have currently been identified. The best studied of these are primary hyperoxaluria type 1 (PH1) and primary hyperoxaluria type 2 (PH2) (reviewed in Danpure and Purdue, 1995; Danpure and Rumsby, 2004; Danpure, 2005). A third hyperoxaluria, classified as "atypical" has been identified but its underlying cause is not understood (Monico *et al.*, 2002). While the enzyme and metabolism specifically involved in PH2 is the basis of this thesis, PH1 will briefly be introduced due to its causative defect being in the same metabolic pathway as PH2 and the clinical similarities of the two disorders.

1.3.1 Primary hyperoxaluria type 1

Primary hyperoxaluria type 1 is a rare, recessive disordered that is caused by mutations within the liver specific peroxisomal enzyme alanine:glyoxylate aminotransferase (AGT, Kamoda *et al.*, 1980; Danpure and Jennings, 1986). AGT catalyses a transamination reaction between the reactive metabolite glyoxylate and alanine, producing pyruvate and glycine (Figure 1.1). When AGT is dysfunctional the build up of glyoxylate is instead removed via its oxidation to oxalate or reduction to glycolate, both reactions being catalysed by glyoxylate oxidase (GO) (Richardson and Tolbert, 1961). Consequently both hyperoxaluria (always) and hyperglycolic aciduria (usually) are signatures of PH1, with the insolubility of calcium oxalate leading to the pathology associated with this disorder.

Over 50 mutations of the AGT gene have been identified with approximately half of these being missense mutations and result in the change of one amino acid for another (Danpure, 2005). The determination of a crystal structure of the AGT enzyme (Zhang *et al.*, 2003) has greatly assisted understanding of these mutations and initiated drug design strategies.

1.3.2 Primary hyperoxaluria type 2

Primary hyperoxaluria type 2 has long been identified as the result of a deficiency in hydroxypyruvate reductase activity, due to its characteristic of both hyperoxaluria and L-glyceric aciduria (Williams and Smith, 1968b). Subsequent identification of a dual lack of both hydroxypyruvate and glyoxylate reductase activity in the liver of PH2 patients (Mistry, Danpure and Chalmers, 1988) has facilitated the identification of the underlying genetic defect; mutations of the glyoxylate reductase/hydroxypyruvate reductase (GRHPR) gene (Cramer *et al.*, 1999; Rumsby and Cregeen, 1999). The resulting metabolic blockage of both glyoxylate and hydroxypyruvate reduction can, as previously mentioned (section 1.2), be overcome by the catalytic ability of the cytoplasmic enzyme L-LDH (Figure 1.1). However, with the majority of glyoxylate now being metabolised by this enzyme, excessive oxalate is produced and results in the pathology seen with PH2. Additionally, the alternative reduction of hydroxypyruvate produces L-glycerate which is not normally seen in urine (Williams and Smith, 1968a), resulting in the defining feature of PH2, L-glyceric aciduria.

PH2 is both less common and less severe then PH1, although there is evidence for its misdiagnosis (Leumann and Hoppe, 1999; Rumsby *et al.*, 2001). The fact that PH2 is less severe then PH1 may arise from glyoxylate itself crossing the peroxisomal membrane and hence being available to AGT without the need for the glycolate intermediate, although the permeability of the peroxisome to glyoxylate has not been determined. Additionally the presence of AGT2 within the mitochondria (Baker *et al.*, 2004) may help to reduce the effect of PH2.

Unlike PH1, only a small number of mutations have been identified within the human GRHPR gene of PH2 patients and have been shown, through expression and activity trials, to be the genetic cause of the disorder (Cramer *et al.*, 1999; Webster *et al.*, 2000; Cregeen *et al.*, 2003; Takayama *et al.*, 2007 and reviewed in Danpure and Rumsby, 2004). The majority of these mutations are small deletions or single base substitutions that result in a truncated peptide due to the creation of termination codons. The most common PH2 mutation is the deletion of guanosine at base 103 (103delG) which is responsible for 39% of known mutated alleles and results in termination of translation after only 44 amino acids (Webster *et al.*, 2000; Cregeen *et al.*, 2003). Further mutations that cause premature termination are scattered throughout the gene sequence, generating another approximately 43% of the identified mutant alleles (Webster *et al.*, 2000; Cregeen *et al.*, 2003). Further identified mutations include missplicing events and 7 missense mutations: Gly165Asp and Met322Arg (Webster *et al.*, 2000), Asn312Asp (personal communication Dr D. Dinour, Israel), Glu113Lys (Takayama *et al.*, 2007), Gly160Arg and Arg302His (personal communication Dr G. Rumsby, UCL Hospitals, London) and Arg302Cys (Cregeen *et al.*, 2003). The majority of these mutations occur in residues conserved within the D-2-hydroxyacid dehydrogenase family or with the GRHPR sub-family (Figure 1.2). As has been demonstrated for the AGT enzyme, the elucidation of the human GRHPR crystal structure will enable the mechanism by which these substitution mutations affect the catalytic ability to be interpreted and should facilitate future drug discovery work. Accordingly, these mutations are discussed in detail in relation to the structure solution presented in Chapter 3.

1.3.3 Diagnosis

While the detection of excessive oxalate in the urine is used as an initial diagnosis tool for suspected primary hyperoxaluria, it does not distinguish between PH1 and PH2 which is important for the determination of the correct treatment course. Additionally, both secondary features, hyperglycolic aciduria (PH1) and L-glyceric aciduria (PH2) are not always present, leading to a suspicion PH2 may be under diagnosed (Danpure and Purdue, 1995; Rumsby *et al.*, 2001).

The identification of the exact enzymatic cause of these two disorders has allowed the introduction of enzymatic testing via liver biopsies, although this is rapidly being superceded by other methods. The identification of the genetic origin and of the most prevalent mutations has allowed the development of a genetic screening approach, although in many cases confirmation by liver analysis is still required (Rumsby, Williams and Coulter-Mackie, 2004). In the case of PH2, as GRHPR is expressed throughout the body (although predominantly in the liver, Cramer *et al.*, 1999) enzymology tests are being extended to blood mononuclear cells (Knight *et al.*, 2006a) although questions have been raised regarding the appropriateness of this method (Bhat, Williams and Rumsby, 2005; Rumsby, 2006).

1.3.4 Treatment options

The treatment options can be divided into those that treat the symptoms of calcium oxalate deposition and those that treat the enzymatic cause (reviewed in Danpure and Purdue, 1995; Danpure and Rumsby, 2004; Danpure, 2005). While the symptoms of PH1 and PH2 are similar, with varying degrees of severity, effective treatment of the underlying cause is dependent on the primary hyperoxaluria type present, making correct diagnosis essential.

Standard treatment of the symptoms during the early stages include: a high fluid intake to increase the excretion of oxalate, restricting the dietary intake of oxalate, and dialysis in later stages. Renal transplantation is also used in late stage cases, but is only a temporary solution as oxalate deposition continues (Danpure and Purdue, 1995). Optimally, treatment should 'fix' the underlying defect. This is achieved in PH1 via liver transplantation as this is the sole location of this enzyme, but has not been used in

PH2 as it is spread more widely throughout the body. In some cases PH1 can be partly regulated by increasing the intake of Pyridoxine (vitamin B_6), the precursor to the AGT cofactor (Watts *et al.*, 1979). Gene therapy and the use of chemical chaperones to stabilise certain AGT mutations have been proposed as future treatments (Danpure, 2005). Again, the more disseminated location of GRHPR makes current gene therapy less suitable for treatment of PH2, and the lack of a structure of the human GRHPR enzyme has precluded substantial drug discovery work. This is further discussed in relation to the GRHPR structure in Chapter 3.

With the establishment of hydroxyproline as the likely precursor of glyoxylate (and therefore oxalate in the primary hyperoxalurias, Knight *et al.*, 2006b) the inhibition of its transport into the mitochondria (or cell) or its metabolism to glyoxylate has been suggested (Coulter-Mackie, 2006). This would have the advantage of being useful for the treatment of both PH1 and PH2. Additionally, the partial inhibition of human L-LDH may be an option in the treatment of PH2, by allowing the glyoxylate blockage to be overcome by other means, such as increased excretion or AGT(2) function. Although complete inhibition of L-LDH is unlikely to be desirable due to its other cellular functions, it is currently being targeted for other diseases (personal communication Prof. L. Brady), although not specifically for PH2.

1.4 D-2-hydroxyacid dehydrogenases

Recognition of sequence homology between a number of plant and bacterial enzymes lead to the establishment of the D-specific 2-hydroxyacid dehydrogenase family distinct from, and unrelated to, the L-specific family (Grant, 1989; Kochhar *et al.*, 1992b). Both types of enzyme catalyse identical chemical reactions, but produce opposite stereo isomer products. Subsequent sequence comparisons have placed the human GRHPR enzyme within the D-specific 2-hydroxyacid dehydrogenase family.

While formate dehydrogenase (FDH) is an atypical member of the family as its is neither stereo specific nor a hydroxyl acid dehydrogenase, its strong sequence homology with D-2-hydroxyacid dehydrogenases (Vinals, Depiereux and Feytmans, 1993) and its early crystallographic structure determination resulted in it being the first enzyme to give structural insight into this family (Lamzin *et al.*, 1994). Other, D-2-

hydroxyacid dehydrogenases followed including a bacterial (*Hyphomicrobium methylovorum*) GRHPR (Goldberg, Yoshida and Brick, 1994), D-2 hydroxyisocaproate dehydrogenase from *Lactobacillus casei* (D-HicDH, Dengler *et al.*, 1997) and a D-lactate dehydrogenase (D-LDH) from *L. bulgaricus* (Razeto *et al.*, 2002). A structure based sequence alignment of these proteins is shown in Figure 1.3.

```
H_GRHPR         ------------------------------------MRPVRLMKVFVTRRIPAEGRVALARAADCEVEQWDSDEPIPAKELERGVAG--AHGLLCLLSDHVDKRILDAAG-ANLKVI  78
P_GRHPR         -----------------------------MAKPVQIEVWNPNGKYRVVSTKPMPGTRWINLLIEQDCRVEICTEKKTILSVEDILALIGDKCDGVIGQLTEDWGEVLFSALSRAGGKAF  90
B_GRHPR         ----------------------------------------------KKKILITWPLP---EAAMARARESYDVIAHGDDPKITIDEMIETAKS-VDALLITLNEKCRKEVIDRIP-ENIKCI  71
D-LDH           ----------------------------------------------TKIFAYAIREDEKPFLKEWEDAHKDVEVEYTDKLLTPETVALAKG--ADGVVVYQQLDYTAETLQALADNGITKM  73
D-HicDH         ----------------------------------------------MKIIAYGARVDEIQYFKQWAKDTGN-TLEYHTEFLDENTVEWAKG--FDGINSLQTTPYAAGVFEKMHAYGIKFL  72
FDH             MAKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGELGLRKYLESNGHTLVVTSDKDGPDSVFERELVD--ADVVISQPFWPAYLTPERIAKAKNLKLA 118
                                                                                                                .  :                 . .

                                                  #                                                   #    #
H_GRHPR         STMSVGIDHLALDEIKKRGIRVGYTPDVLTDTTAELAVSLLLTTCRRLPEAIEEVKNGGWTSWKPLWLCGYGLTQSTVGIIGLGRIGQAIARRLKP-FGVQRFLYTGRQ-PRPEEAAEFQ 196
P_GRHPR         SNMAVGYNNVDVNAANKYGVAVGNTPGVLTETTAELAASLSLAAARRIVEADEFMRAGRYDGWLPNLFVGNLLKGQTVGVIGAGRIGSAYARMMVEGFKMNLIYFDLYQSTRLEKFVTAY 210
B_GRHPR         STYSIGFDHIDLDACKARGIKVGNAPHGVTVATAEIAMLLLLGSARRAGEGEKMIRTRSWPGWEPLELVGEKLDNKTLGIYGFGSIGQALAKRAQG-FDMDIDYFDTHR-ASSSDEASYQ 189
D-LDH           SLRNVGVDNIDMAKAKELGFQITNVPVYSPNAIAEHAAIQAARILRQDKAMDEKVARHDLR-WAPT--IGREVRDQVVGVIGTGHIGQVFMQIMEG-FGAKVIAYDIFRNPELEKKGYYV 189
D-HicDH         TIRNVGTDNIDMTAMKQYGIRLSNVPAYSPAAIAEFALTDTLYLLRNMGKVQAQLQAGDYE--KAGTFIGKELGQQTVGVMGTGHIGQVAIKLFKG-FGAKVIAYDPYP-MKGDHPDFDY 188
FDH             LTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEWARKGGWN-IADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAP-FDVHLHYTDRHRLPESVEKELNL 236
                    :* ::: :      .. : ..  . : ** .       *.                        .  :    :*  . * ** .  :     *  .             .

H_GRHPR         AEFVSTP--------------ELAAQSDFIVVACSLTPATEGLCNKDFFQKMKETAVFINISRGDVVNQDDLYQALASGKIAAAGLDVTSPEPL--------------PTNHPLLTLKNC 288
P_GRHPR         GEFLKANGEAPVTWRRASSMDEVLREADVISLHPVLDKTTFHLVNKESLKAMKKDAILINCSRGPVIDEAALVDHLRDNPMFRVGLDVFEDEPY--------------MKPG-LADMKNA 315
B_GRHPR         ATFHDSLD-------------SLLSVSQFFSLNAPSTPETRYFFNKATIKSLPQGAIVVNTARGDLVDNELVVAALEAGRLAYAGFDVFAGEPN--------------INEG-YYDLPNT 281
D-LDH           DSLDDLYK-----------------QADVISLHVPDVPANVHMINDESIAKMKQDVVIVNVSRGPLVDTDAVIRGLDSGKIFGYAMDVYEGEVG-IFNEDWEGKEFPDARLADLIARPNV 291
D-HicDH         VSLEDLFK-----------------QSDVIDLHVPGIEQNTHIINEAAFNLMKPGAIVINTARPNLIDTQAMLSNLKSGKLAGVGIDTYEYETEDLLNLAKHG-SFKDPLWDELLGMPNV 290
FDH             TWHATRED--------------MYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNTARGKLCDRDAVARALESGRLAGYAGDVWFPQPA--------------PKDHPWRTMPYN 328
                                          .:.. :      .  : *.  :  :   . .:* :*  : :   :   *  . :   . *.   :

                              #         #         #
H_GRHPR         VILPHIGSATHRTRNTMSLLAANNLLAGLRGEPMPSELKL-------------------------------- 328
P_GRHPR         IIVPHIASASKWTREGMATLAALNVLGKIKGYPVWSDPNRVEPFLDENVSPPAASPSIVNAKALGNA------ 382
B_GRHPR         FLFPHIGSAATQAREDMAHQANDLIDALFGGADMSYALA--------------------------------- 320
D-LDH           LVTPHTAFYTTHAVRNMVVKAFDNNLELVEGKEAETPVKVG------------------------------- 332
D-HicDH         VLSPHIAYYTETAVHNMVYFSLQHLVDFLTKGETSTEVTGPAK----------------------------- 333
FDH             GMTPHISGTTLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSEEAAKFKKAV 401
                 : ** .  :  :       :       .
```

Figure 1.2 Sequence alignment of representative D-2-hydroxyacid dehydrogenases.

H: Human, P: Plant (cucumber, Greenler *et al.*, 1989), B: Bacterial (*H. methylovorum*). Sequences (apart from cucumber) are from structures referenced in the text. Catalytic residues are coloured red. Missense mutations that lead to PH2 are indicated by hashes (green). This alignment will be referred to throughout this thesis and as such residues thought to be involved in substrate specificity (as determined in this thesis and discussed in Chapters 3 and 4) are in blue. Alignment performed in ClustalW (Larkin *et al.*, 2007) with manual structure based adjustments.

Each monomer of these dimeric enzymes consists of two distinct α/β/α globular domains: the coenzyme binding domain and the substrate binding domain (Figure 1.3). Both domains have similarity to the Rossmann fold and share a core structure (Figure 1.3 insert). Between these domains is the active site cleft in which the binding sites for coenzyme and substrate are located. A comparison of apo and holo structures has shown that during catalysis the substrate binding domain moves as a rigid body towards the coenzyme binding domain, leading to shielding of the active site from solvent (Figure 1.3, Lamzin *et al.*, 1994; Razeto *et al.*, 2002).

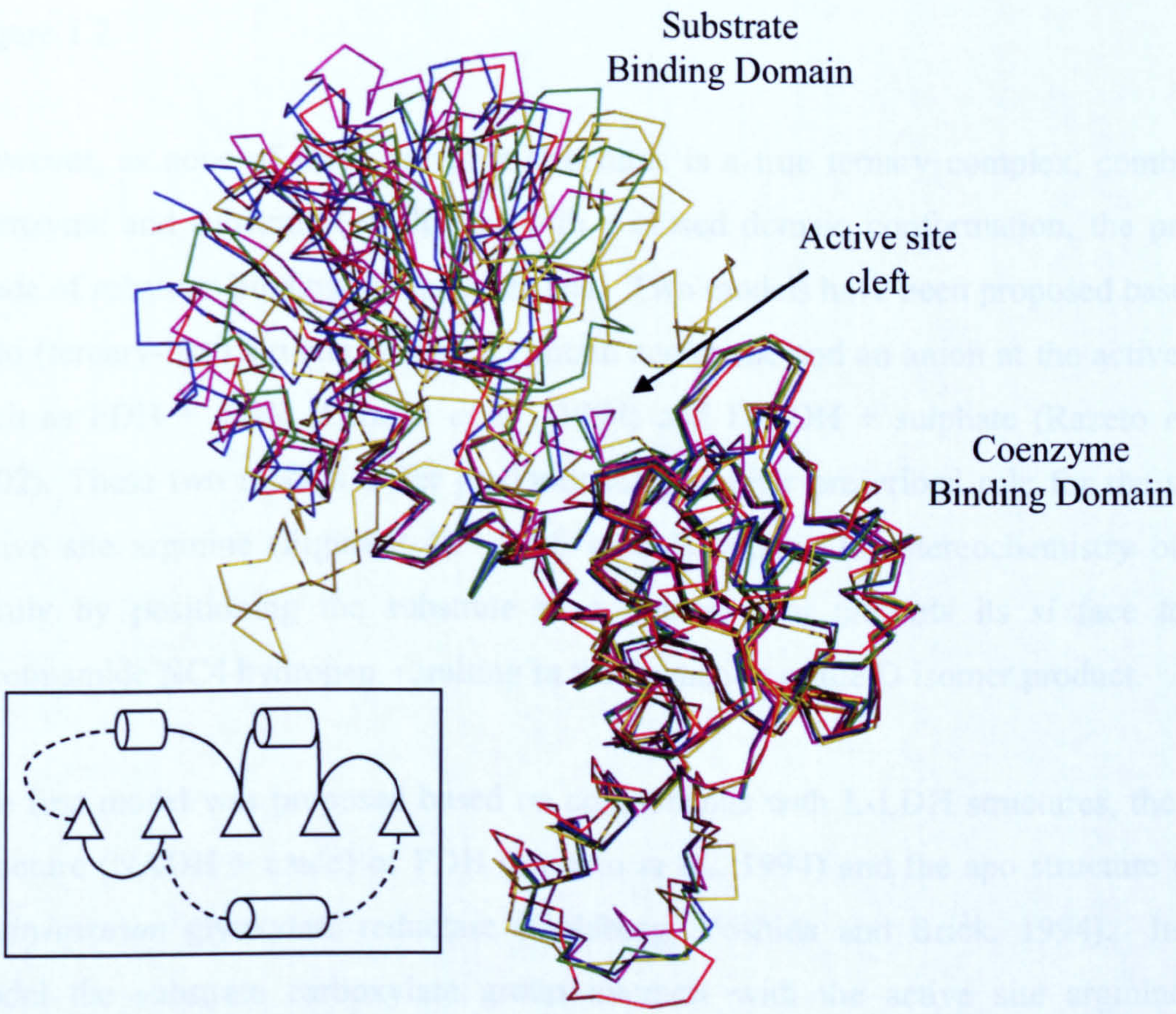


Figure 1.3 Overlay of monomers from characteristic D-2-hydroxyacid dehydrogenases.

A) Overlay of monomers superimposed via their coenzyme binding domain. FDH: holo (yellow, Lamzin *et al.*, 1994), D-LDH: holo (green, Razeto *et al.*, 2002) D-LDH: binary (blue, Razeto *et al.*, 2002), D-GDH: apo (red, Goldbery, Yoshida and Brick, 1994), D-HicDH: ternary (magenta, Dengler *et al.*, 1997). Boxed insert is a schematic of the core domain structure conserved between domains. Cylinders represent α–helices, triangles β-sheets and dashed lines non-direct links (adapted from Kutzenko, Lamzin and Popov, 1998).

1.4.1 Binding of substrate

While a growing number of D-2-hydroxyacid dehydrogenase protein structures have been solved, uncertainty had remained regarding the exact orientation of substrate within the active site. The identities of the residues involved in the catalytic activity of the D-2-hydroxyacid dehydrogenase family have long been established by mutational and chemical modification studies (Kochhar, Chuard and Hottinger, 1992; Kochhar *et al.*, 1992a; Taguchi and Ohta, 1993, 1994). These residues have subsequently been confirmed in corresponding structural work (Goldberg, Yoshida and Brick, 1994; Lamzin *et al.*, 1994; Dengler *et al.*, 1997; Razeto *et al.*, 2002) and are indicated in Figure 1.2.

However, as none of these previous structures is a true ternary complex, combining coenzyme and substrate (-analogue) with a closed domain conformation, the precise mode of substrate binding has been unclear. Two models have been proposed based on holo (ternary-like) structures, which contain coenzyme and an anion at the active site; such as FDH + azide (Lamzin *et al.*, 1994) and D-LDH + sulphate (Razeto *et al.*, 2002). These two models differ predominantly in their prescribed role for the single active site arginine (Figure 1.4). Both models explain the stereochemistry of this family by positioning the substrate in a manner that presents its *si* face to the nicotinamide NC4 hydrogen, resulting in the formation of the D-isomer product.

The first model was proposed based on comparisons with L-LDH structures, the holo structure (NADH + azide) of FDH (Lamzin *et al.*, 1994) and the apo structure of *H. methylovorum* glyoxylate reductase (Goldberg, Yoshida and Brick, 1994). In this model the substrate carboxylate group interacts with the active site arginine via bifurcated hydrogen bonds (Figure 1.4), similar to the interaction seen in L-LDH (Wigley *et al.*, 1992).

A second binding mode was later proposed, initially based on a structural model of *L. bulgaricus* D-LDH (Vinals *et al.*, 1995) and again when the crystal structure of D-LDH from *L. pentosus* was found to closely match this (Stoll, Kimber and Pai, 1996). In this model the active site arginine forms bifurcated charged hydrogen bonds to a single

carboxylate oxygen of the substrate and also its keto oxygen (Figure 1.4). In this model the arginine both acts in substrate binding and catalysis via the polarisation of the substrate carbonyl bond, which is supported by kinetic studies (Taguchi and Ohta, 1994). While partial structural support for this second model is available, it is based on structures with inappropriate domain conformations (Dengler *et al.*, 1997) or unusual active site features (Kumar *et al.*, 2002). Consequently, despite this second model of substrate binding being generally accepted (Figure 1.4b), no structural confirmation has previously been produced.

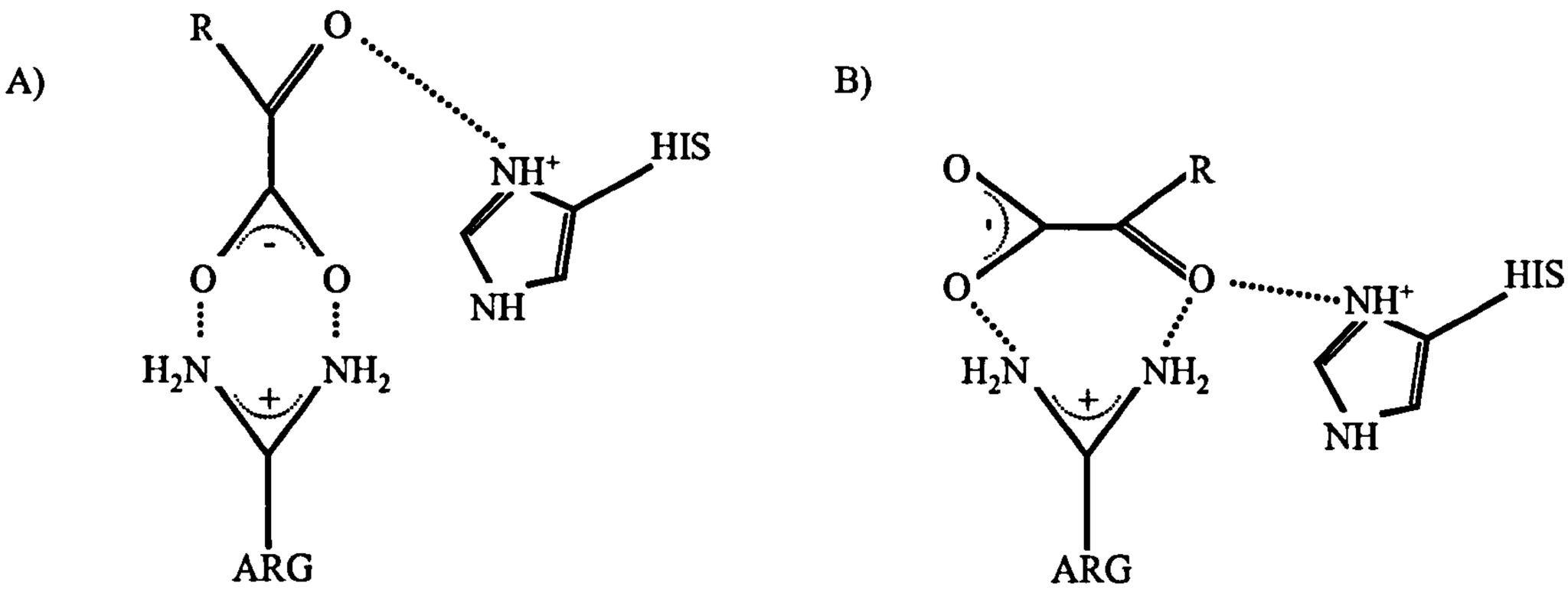


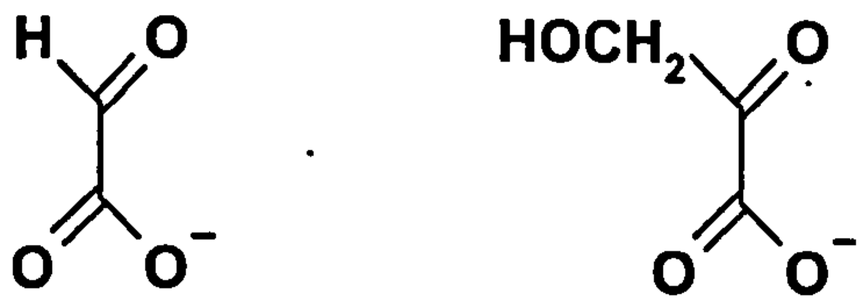
Figure 1.4 The two models for substrate binding in the D-2-hydroxyacid dehydrogenase family.

A) The active site arginine forms bifurcated bonds to the carboxylate group of the substrate ('R' represents any attached group). B) The arginine binds the carboxylate group and polarises the carbonyl bond, assisting catalysis. In model (b) additional residues are involved in binding the carboxylate group but are not included for clarity. Adapted from Razeto *et al.*, (2002).

1.4.1.1 Substrate specificity

Substrate specificity varies widely across the D-2-hydroxyacid dehydrogenase family. D-HicDH has low specificity and will act on a large range of 2-oxocarboxylic acids with various side groups (represented by 'R' in Figure 1.4, Dengler *et al.*, 1997). In contrast, the GRHPR members are highly specific and will only catalyse the substrates glyoxylate and hydroxypyruvate (Figure 1.5), which have very specific groups. While these same substrates are present in plants (Givan and Kleczkowski, 1992), bacteria (Chistoserdova and Lidstrom, 1991) and humans (Rumsby and Cregeen, 1999) the structural determinants of this substrate specificity are unknown. While one of the earliest D-2-hydroxyacid dehydrogenase structures solved was that of a bacterial

GRHPR (Goldberg, Yoshida and Brick, 1994) this provided minimal information on the mechanism of substrate specificity in this family as it was in the apo form with nothing bound in the active site. Since the solution of the human ternary GRHPR structure presented in this thesis, the structure of two further bacterial GRHPR enzymes have been produced from *Thermus thermophilus* (PDB code 2CUK) and *Pyrococcus horikoshii* (Yoshikawa *et al.*, 2007). Both of these enzymes show a high level of sequence identity to the human form, 42% and 39% respectively. However, as these enzymes have only been solved either in the apo state (*T. thermophilus*) or with sulphate in the active site (*P. horikoshii*) they do not assit in the understanding of substrate specificity of the group. The determinants of this specificity can only be reliably understood with the structure of one of these GRHPR enzymes with a substrate analog bound in the active site.



glyoxylate **hydroxypyruvate**

Figure 1.5 Schematic of the two physiological substrates of human GRHPR.

1.5 Thesis aims

To further understanding of the GRHPR family of enzymes and their role in PH2 disease, studies described in this thesis aimed to address two main goals:

1) To produce crystallographic structures of the three states of human GRHPR (apo, binary and ternary) and use these to understand its substrate and coenzyme specificity and the pathology of PH2 mutations. Additionally, structural confirmation of the mode of substrate binding within the D-2-hydroxyacid dehydrogenase family was to be clarified (Chapter 3).

2) To characterise the kinetics of human GRHPR under 'near-physiological' conditions in order to allow their direct comparison to similarly produced

results from L-LDH. It was intended that this information could be used to help understand the flux of these metabolites through the glyoxylate metabolic pathway (Chapter 4).

Additionally, a number of other projects involving unrelated components were also initiated. These will be introduced later in the thesis, and aimed to:

3) Establish soluble expression and purification procedures for human Sorting Nexin 1 (SNX1), or parts thereof, to initiate structural studies of its ligand specificity (Chapter 5); and

4) Using crystallography, to investigate the interaction of two plant RKIP/PEBP proteins, TFL1 and FT, with their potential phosphorylated ligands (Chapter 6).

Chapter 2 Methods

2.1 Methods overview

This chapter consists of all the methods specific to the work undertaken during this thesis on human GRHPR. In addition, a number of the methods described herein were used in relation to the two additional projects described in Chapter 5 and Chapter 6 and will be referred back to during those chapters where appropriate.

2.2 General materials

All buffers were made from high grade products and using Milli Q water (Millipore). Large volumes (>50ml) were vacuum filtered through 0.2µm cellulose nitrate filters (Sartorius), while smaller volumes were filtered using 0.2µm cellulose acetate syringe filters (Sartorius). Size exclusion chromatography (gel filtration) buffers were degassed before use and unless indicated, all columns were run using an AKTA purifier controlled by the Unicorn software (GE Healthcare). To prevent cross-contamination all Histidine (poly-His) tagged proteins were purified on individual HiTrap Ni^{2+} affinity columns (Amersham Biosciences).

2.3 GRHPR

2.3.1 Supplied GRHPR

Human GRHPR cDNA was supplied by Dr G. Rumsby (University College London Hospitals, London) pre-cloned into the expression vector pTrcHisB. This construct consists of the 987 base pair (bp) open reading frame coding for the 328 amino acid native protein plus a small length of 5' and 3' untranslated sequence. This had been inserted in-frame and downstream from a 6 histidine (hexa-His) tag via *Kpn*I and *Hind*III restriction sites (5' and 3' respectively). Expression from this results in a 40 kDa N-terminally His-tagged protein, comprising a 48 amino acid tag and 328 amino acids of native GRHPR. Full cloning and initial expression/activity tests can be found in Rumsby and Cregeen (1999). The vector also contains the ampicillin resistance gene

and for selection purposes 50μg.ml^{-1} of ampicillin was always present. This construct is referred to as pTrcHisB-HPR from here on.

2.3.2 PCR modification of GRHPR

The pTrcHisB-HPR vector was modified by the polymerase chain reaction (PCR) to include an N-terminal fused thrombin cleavage sequence (LVPRAS) directly before the start of the native enzyme. This would allow for the removal of all but two of the tag amino acids (AS). A forward primer was designed containing a *Kpn*I restriction site, the thrombin recognition sequence and 25bp complementary to the start of the open reading frame, while the reverse primer was from Rumsby and Cregeen (1999) and contains a *Hind*III restriction site (Figure 2.1). All primers used in this thesis were produced by Invitrogen and supplied as lyophilised powder. The primers were resuspended in Elgastat water to a final concentration of 0.1mM and stored at -20°C until required.

Forward Primer:

5'-CGGGTACC**CTGGTGCCGCGCGCCAGC**ATGAGACCGGTGCGACTCATGAAGG-3'

Reverse Primer:

5'-GCAAGCTTCCCTTGGCTCTGCCTGTGGATCC-3'

Figure 2.1 Primers for thrombin site insertion into GRHPR.

Restriction enzymes sites are underlined, with a *Kpn*I site in the forward primer and a *Hind*III site in the reverse. The DNA sequence coding for the thrombin recognition sequence (LVPRAS) is in bold.

The PCR reaction mix consisted of 1.75 enzyme units (U) Expand High Fidelity Polymerase (Roche), 1x Expand Buffer + $MgCl_2$ (Roche), both primers at 2μM, dNTPs (50pmol) and 1μl of pTrcHisB-HPR as the template. This was made up to a final volume of 50μl with Elgastat water. A PCR machine (Progene) was used to create the thermal cycling program required for amplification. Initial DNA denaturation was undertaken at 94°C for 4 minutes followed by 25 cycles of 94°C for 1 minute, 48°C for 45 seconds, 72°C for 1 minute for denaturation, annealing and amplification respectively. PCR reactions were then held at 4°C until required.

2.3.3 PCR product verification and purification

To ascertain the success of the PCR, Tris-HCl-acetate-EDTA (TAE) buffer was used to prepare a 1% agarose gel. Both the gel and running buffer (TAE) contained a small amount of ethidium bromide (EtBr), which upon intercalation with double stranded DNA and exposure to UV light fluoresces, provided a means of visualising the DNA. The PCR reactions were electrophoresed through this gel for 45min at 110 volts. Correctly sized DNA products were identified by comparison to concurrently run molecular weight markers (Hyperladder, Bioline) under UV light. Care was taken to minimise sample exposure to UV light as this can cause mutations within the PCR product and inhibit later ligations. PCR products of the correct size were sliced from the gel using a sterile scalpel and subsequently extracted from the agarose using the Quick Gel Extraction Kit (Qiagen), following manufacturers' instructions. Purified PCR product could be stored at -20°C indefinitely.

2.3.4 pGEM-Teasy ligation

Many restriction enzymes are known to cleave inefficiently when their recognition sequence is near the end of a linear DNA fragment, making it difficult to clone PCR products directly into expression vectors (Marchuk *et al.*, 1991). This problem was avoided by using an intermediate cloning step into the pGEM-Teasy vector system (Promega). This linearised, high-copy number vector contains single 3' terminal thymidines at both ends. These thymidines prevent empty vector recircularisation and provide overhangs directly compatible with the extra deoxyadenosine added to the 3'-ends of PCR products by the Expand High Fidelity Polymerase. Additionally, PCR insertion takes place within the coding region of the α-peptide of β-galactosidase, allowing the identification of successful recombinants via blue/white screening. This vector contains the ampicillin resistance gene and unless stated 50μg.ml^{-1} of ampicillin was present in all cultures.

Ligations into pGEM-Teasy were performed in accordance with the manufacturers protocol. Briefly, 5μl of 2x Rapid Ligation Buffer, 1μl of linear pGEM-Teasy (50ng) and 1μl T4 DNA Ligase (3U) were mixed with 3μl of purified PCR product. A negative control ligation was also run, where the PCR product was replaced with

Elgastat water. Ligation reactions occurred at 4°C overnight and were then frozen at -20°C until transformed.

2.3.5 Bacterial transformation

To ensure efficient transformation and allow for blue/white selection, *Escherichia coli* NovaBlue Singles™ super calcium competent cells (Novagen) were used. For each transformation, a 25µl aliquot of cells was allowed to thaw on ice and a 1.5µl sample of the ligation mix added. A 5 minute incubation on ice was followed by a 40 second, 42°C heat shock in a thermostated water bath. This induces plasmid uptake by the bacteria, which are then placed back on ice for 2 minutes. Room temperature SOC media was added to give a final volume of 100µl. An aliquot of the transformed bacteria (30µl) were spread onto pre-treated Luria Bertani (LB) agar plates containing ampicillin. Pre-treatment of the LB agar plates consisted of them being spread with 20µl of 0.1M Isopropyl-β-D-thiogalactopyranoside (IPTG) and 60µl of 20mg.ml^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), then left at 37°C to fully dry. This provides the induction and substrate required for blue/white selection. Plates were incubated at 37°C overnight.

2.3.6 Blue/white selection

Blue/white colony screening allows for the easy choice of likely successfully ligated plasmids. It relies on the fact that the *E. coli* β-galactosidase enzyme (LacZ) will turn the normally colourless X-gal into an insoluble blue product. The addition of the IPTG induces the NovaBlue strain to express the LacZΩ peptide, while only recircularised but empty pGEM-Teasy will express active LacZα peptide. Both active peptides are required for full LacZ activity and result in blue colonies. The presence of a ligated insert in pGEM-Teasy inactivates the LacZα peptide, rendering these colonies white.

After overnight growth at 37°C, transformation plates were examined for white colonies. Six white colonies were picked with sterile toothpicks, re-streaked onto ampicillin/LB agar plates and re-grown overnight at 37°C.

2.3.7 Verification of inserted DNA and sequencing

A small number of cells from each of the 6 selected colonies were transferred to individual universals containing 10ml LB broth and ampicillin. These were grown with shaking (300 rpm) overnight at 37°C. Plasmid DNA was extracted from 5ml of these cultures using the QIAprep Spin Miniprep kit (Qiagen), following manufacturers instructions. Plasmid DNA was eluted from the spin columns in 50µl of storage buffer (10mM Tris-HCl, pH 8.5). To verify the presence of insert DNA (blue/white selection is not 100% accurate) double restriction digests of the purified plasmid were performed. In brief, a 5µl aliquot of purified plasmid was digested with 1µl each of *Kpn*I and *Hind*III (all restriction enzyme used in this thesis were supplied by Roche) in total volume of 25µl. Digestion reactions were incubated at 37°C for a minimum of an hour and analysed by 1% agarose gel electrophoresis. Correctly sized inserts were identified by comparison with known molecular weight markers under UV light.

Two plasmids containing correctly sized inserts were sent for gene sequencing (Cogenics Inc.) using the pGEM primers M13 forward and M13 reverse, which flank the insert. Only plasmids of confirmed, correct sequence were then used for further sub-cloning.

2.3.8 Expression vector preparation

The pTrcHisB-HPR construct was digested to remove the original GRHPR coding region and prepare it for the insertion of the modified version. This was done by *Kpn*I/*Hind*III double digestion on a 15µl aliquot of pTrcHisB-HPR in a total volume of 20µl. Digestion occurred over an hour at 37°C, after which calf alkaline phosphatase (CAP) was added to catalyse the removal of 5′ phosphate groups further preventing recircularisation. Digested and phosphatased vector was electrophoresed through a 1% agarose gel and a linear, vector sized band was identified, excised from the gel and purified away for the gel by the Quick Gel Extraction Kit. To provide insert, pGEM-Teasy containing the PCR product was treated in the same fashion, leaving out the phosphatase step and gel purifying the insert sized band. Care was taken to avoid excessive UV exposure and excise only the required DNA band.

Ligation and transformation was implemented as described for pGEM-Teasy (sections 2.3.4 and 2.3.5). Briefly, linearised pTrcHisB vector was mixed with digested insert in a 1:6 ratio (0.5µl:3µl) and T4 ligase (1µl). After overnight incubation at 4°C, 2µl of ligation mix was transformed into the *E. coli* expression strain Rosetta (Novagen). Due to resistance of this strain to chloramphenicol these transformants were given 30 minutes at 37°C to recover in the absence of antibiotics. The entire transformed cell mix (100µl) was then spread on LB agar plates containing ampicillin and chloramphenicol (34µg.ml^{-1}). Large colonies present after overnight growth at 37°C were selected and the presence of the insert confirmed by double digestion as described in section 2.3.7. The thrombin cleavable construct will be referred to as pTrcHisB-T-GRHPR from here on and its expression product as T-GRHPR.

2.3.9 Cell line storage

For short term storage and direct use, transformed bacterial cell lines were streaked onto LB/agar plates containing appropriate antibiotics. These plates provided a source of individual colonies and were stored at 4°C for a maximum of one week, before being re-streaked onto fresh plates. Cells for long term storage were first grown as 10ml LB broth cultures at 37°C overnight. A 0.7ml aliquot of these cells was mixed with 0.3ml of 50% sterile glycerol and stored at -80°C.

2.3.10 Expression tests and scale up

A single colony was picked using a sterile toothpick and used to inoculate a 10ml LB broth culture containing appropriate antibiotics. After overnight grown at 37°C with shaking (300rmp), 1ml of this was used to inoculate a 100ml LB culture. Growth of this culture at 37°C was monitored by optical density (OD) measured at 600nm (Lambda 25 spectrophotometer, Perkin Elmer). Upon reaching an OD of approximately 0.6, a 5ml sample was taken (uninduced) and recombinant protein expression induced in the remainder by the addition of IPTG to a final concentration of 1mM. Induced 5ml samples were taken at 1 hour time points for the following 5 hours and again after overnight expression. All samples were centrifuged (5,000g, 10 minutes), the supernatant discarded and the pellet frozen at -20°C.

Each cell pellet was resuspended in 500µl phosphate buffered saline (PBS) and sonicated (VibraCell sonicator, Sonics) using a micro tip. Sonication was performed at an amplitude (Ampl) of 35% for 3 x 15 second bursts with samples kept on ice throughout. A 10µl sample of total protein was taken prior to the remainder being centrifuged (18,000g, 10 minutes) to separate soluble from insoluble, with a subsequent 20µl sample of the soluble supernatant being taken. Recombinant protein expression was analysed on 12% Sodium dodecyl sulphate Polyacryliamide Gels (SDS-PAGE). Both samples had SDS sample loading buffer (30% glycerol, 5% SDS, 50mM EDTA, 50mM Tris-HCl pH 7.0, 5% β-mercaptoethanol) added (3:1 ratio) and were heated to 100°C for 5 minutes before SDS gel loading. Electrophoresis was run at 180V for 60 minutes, after which gels were coomassie stained (25% ethanol, 10% acetic acid, 0.25% w/v Coomassie Blue) and destained (20% ethanol, 5% acetic acid) to allow visualisation of protein bands. Recombinant protein was identified by comparison to a simultaneously run molecular weight marker (Sigma). All protein gels were undertaken in an identical manner.

Large scale expression proceeded in 1L LB media and was largely the same as for trial expressions, except the entire 10ml overnight culture was used to inoculate the litre. After protein expression had been induced for the appropriate length of time, bacterial cells were pelleted by centrifugation (5,000g, 5 min, Sorvall RC6), resuspended in 40ml PBS and re-centrifuged (5,000g, 10 min, Sorvall Legend RT). These cell pellets could be stored at -20°C until required.

2.3.11 Nickel affinity purification

The expression of recombinant protein with a hexa-His tag allows for a quick and easy first purification step due to the tight, reversible affinity of the tag for nickel (Ni^{2+}) (Schmitt, Hess and Stunnenberg, 1993). To purify recombinant T-GRHPR, a large scale expression cell pellet was resuspended in 30ml GR Nickel Loading Buffer (20mM Tris-HCl pH 8.5, 500mM NaCl, 20mM imidazole, 10% glycerol, 1mM β-mercaptoethanol) with an EDTA-free protease inhibitor (Roche). Cell disruption was accomplished by sonicating the resuspension on ice at 60% Ampl, for 5 x 20 second bursts using the standard probe (VibraCell, Sonics). A 20 second pause between each burst was used to prevent sample overheating and potential protein denaturation. Cell

debris and insoluble proteins were pelleted by centrifugation at 15,000g for 25 minutes (Sorvall RC6). A 5ml HiTrap Ni^{2+} affinity column (Amersham Biosciences) was freshly charged with Ni^{2+} and equilibrated in buffer, according to manufacturers instructions. The clarified supernatant was loaded onto this column manually at approximately $2ml.min^{-1}$ using a 20ml syringe. The flow through was retained and the column attached to an AKTA purifier. At a flow rate of $1ml.min^{-1}$, 5 - 10 column volumes (CV) of Loading Buffer (or until a flat A_{280nm} reading was achieved) was used to wash non-specifically bound protein from the column. Bound protein was then eluted over a 0.02M – 1M gradient of imidazole at a flow rate of $1ml.min^{-1}$ and 1ml fractions collected. Fractions corresponding to A_{280nm} peaks were kept, analysed by SDS-PAGE and those containing recombinant protein pooled.

2.3.12 Measuring protein concentration

The amount of protein present in the pooled sample was measured by absorbance at a wavelength of 280nm (Lambda 25 spectrophotometer, Perkin Elmer). Because the low grade imidazole used in the purification contains contaminants that also absorb at this wavelength, an appropriate blank was always used. The predicted molecular weight of T-GRHPR (40,651 Da) and its theoretical extinction coefficient (29450 $M^{-1}cm^{-1}$) were both calculated with Protparam (Expasy; http://ca.expasy.org) and used to estimate protein concentration using the Bear-Lambert equation ($A = \varepsilon.l.c$).

2.3.13 Removal of the poly-His tag

To facilitate removal of the tag, 10U of thrombin (Amersham Biosciences) were added per mg of recombinant protein. The mixture was left for 16hrs at 18°C to ensure complete cleavage. The digested protein was then buffer exchanged into GR Nickel Loading Buffer by rounds of dilution and subsequent concentration using a 10kDa cut-off Vivaspin concentrator (Vivascience). This was passed through a HiTrap Ni^{2+} affinity column connected directly to a 1ml Benzamidine column (Amersham Biosciences). The Ni^{2+} column bound any uncleaved protein (T-GRHPR) and free tag, while the Benzamidine column removed the thrombin. The flow through contained cleaved GRHPR and was retained for further purification.

2.3.14 Gel filtration

The resulting GRHPR was buffer exchanged into GR Gel Fil. Buffer (20mM Tris-HCl pH 8.5, 100mM NaCl, 1mM β-mercaptoethanol) and concentrated (Vivaspin 10 kDa MWCO, Vivascience) to approximately 0.5ml. Gel filtration was run on a pre-equilibrated HiLoad 16/60, Superdex 75 prep grade column (Amersham Biosciences) run at 0.3ml.min^{-1} and 2ml fractions were collected. Those corresponding to A_{280nm} peaks were kept and analysed by SDS-PAGE, with those containing pure recombinant protein pooled and used in all further studies. The predicted molecular weight of GRHPR (35,826 Da) and its theoretical extinction coefficient (27960 $M^{-1}cm^{-1}$) were used for protein concentration calculations.

2.3.15 GRHPR crystallisation

Purified GRHPR was buffer exchanged into 20mM Tris-HCl pH 8.5 with 1mM β-mercaptoethanol by 3 rounds of 20-fold dilution and concentration (Vivaspin 10 kDa MWCO, Vivascience). A final protein concentration of 5.5 mg.ml^{-1}, as calculated by A_{280} readings was achieved and used in crystallisation trials. All crystallisations were performed by the vapour-diffusion method.

2.3.15.1 Initial crystallisation screens

GRHPR, like other D-2-hydroxyacid dehydrogenases was expected to be able to form 3 different enzyme:ligand states; apo – no ligands, binary – just co-enzyme bound and ternary – co-enzyme and "substrate-like" molecule. Crystallisation trials were implemented in attempts to crystallise all 3 states. Concentrated protein was incubated on ice for 30 min with either just nicotinamide adenine dinucleotide phosphate (NADPH, Sigma) or NADPH and di-sodium oxalate (Sigma). Additionally, two different concentrations of these were tried. Table 2.1a lists the different protein:ligand combinations set up in attempts to produce the 3 different states. These incubations were then centrifuged (18,000g, 5min) prior to the crystal tray set up. Initial crystal screens were established as sitting drop vapour-diffusions in 96-well plate format (Intelli Plate, Hampton Research) using a number of commercially available crystal screens (Table 2.1b).

a)

Crystal State	Co-enzyme	"Substrate-like"
Apo	-	-
Binary	2mM NADPH	-
Binary	0.2mM NADPH	-
Ternary	2mM NADPH	5mM di-sodium oxalate
Ternary	0.2mM NADPH	0.5mM di-sodium oxalate

b)

Crystal Screen	Manufacturer	No. of conditions
Crystal Screen (CS) I and II	Hampton Research	96
PEG/Ion Screen	Hampton Research	48
CS Lite	Hampton Research	48
JBScreen 1 – 4	Jena Bioscience	96
JBScreen 5 – 8	Jena Bioscience	96
JBScreen 9 and 10	Jena Bioscience	48

Table 2.1 Crystallisation screening for three GRHPR states.

a) protein:ligand combinations set up during the screening process, b) commercial screens used for initial condition screening.

To set up the screens, 8-channel pipettes were used to allow entire columns to be done simultaneously. This minimised the length of time each drop was exposed to the air and therefore decreased evaporation. Firstly, 50µl of screen solution mother liquor was transferred to the reservoirs. Then using an electronic 8-channel pipette, a 1µl air gap was aspirated, followed by 2µl of protein 'complex' and finally 2µl of reservoir solution. This was then dispensed in to the upper sample well of the crystallisation plate, with the air gap ensuring all liquid was expelled. No manual mixing of the drops was performed. The plates were sealed (ClearSeal Film, Hampton Research) and left to equilibrate at 18°C for at least 48hrs before being inspected for crystal growth with a light microscope (Wild MZ8, Leica). The plates were subsequently re-checked regularly.

2.3.15.2 Crystal optimisation

Screen conditions containing successful 'hits' – defined as the presence of birefringent 'crystals' or phase separation – were selected for further screening in an attempt to produce improved crystals. Screen conditions were used as a starting point around which the precipitant concentration and buffer pH was varied. During optimisation, the buffer pH was altered by 0.5 (initial) and 0.25 (fine screening) pH units, salts were varied in 0.05M steps and polyethylenglycols (PEGs) in 2% w/v alterations. Optimisation was attempted in VDX™ pre-greased 24-well plates (Hampton Research), using the hanging drop vapour diffusion method. To begin with, 700µl of optimisation mother liquor was transferred to the reservoir, with 2µl pipetted onto the centre of an OptiClear Plastic™ cover slip (Hampton Research). Next, 2µl of the protein 'complex' was pipetted directly onto the mother liquor drop, the cover slip carefully inverted and sealed over its corresponding reservoir. As with the screens, these plates were left to equilibrate at 18°C for at least 48hrs before being inspected for crystal growth.

Further improvement of selected conditions was attempted using the Additive Screen (Hampton Research). Here, a 24-well plate is used to screen a single mother liquor condition using the addition of a small amount of various ions, salts, organics and chaotropes. This was set up following manufactures' instructions, with identical reservoir solutions and 2µl protein/2µl reservoir/0.2µl additive drops. Volatile additives were added in a 1:10 ratio to the reservoir solution and drops set up directly from this.

2.3.15.3 Seeding

Both seeding and streaking (below) are a method of introducing nucleation points into a crystallisation condition that may not produce them on it own. It provides a mechanism to separate favourable crystal growth conditions from favourable crystal nucleation conditions, and can dramatically improve crystals. These techniques were only applied during apo protein crystallisation.

Crystal seeds were created using the Seed Bead Kit (Hampton Research) following the manufacturers' instructions. Briefly, the Seed Bead Kit consists of a microcentrifuge

tube containing a small ball, the seed bead. As the apo crystals were quite small, a number of them were selected and transferred to a large drop (5µl) of their mother liquor. This was pipetted into the seed bead tube with another 40µl of their mother liquor. Where the drop had been was rinsed with a further 5µl to make sure all crystals had been transferred, giving a total final volume of 50µl. The tube was then vortexed for 90 seconds, allowing the seed bead to crush the crystals. This provided the "seed stock" and was further diluted with mother liquor from the original condition as needed. Seeding was initially undertaken using the Phoenix nanolitre robot (Art Robbins Instruments) into Structure Screen I and II HT-96 and PACT *premier*™ HT-96 screens (Molecular Dimensions Limited, MDL). This was based on the robotic seeding method of D'Arcy, Villard and Marsh, (2007). Seed stock dilutions of 10- and 200-fold were used. The screen was set up at with 3.5mg.ml^{-1} of GRHPR, using a 1:2:3 ratio of seed solution, reservoir and protein (0.1µl:0.2µl:0.3µl). The 96-head first deposits 50µl reservoir solution then the 0.2µl sitting drop. This is followed by the nano-head adding the protein (0.3µl) then the seed solution (0.1µl), with extensive washing between. Promising hits were then further screened by manually setting up hanging drops with the same 1:2:3 ratio (0.6µl:1.2µl:1.8µl) and seed dilutions of 10- and 50-fold.

2.3.15.4 Streaking

Streaking is attempted where supersaturation of the drop has occurred but spontaneous nucleation cannot. This is known as the metastable zone. To find this zone, the condition that produced the best apo crystal was selected and re-screened over a stepwise reduction in protein or precipitant concentration. Each new condition was prepared as triplicate hanging drops to allow for seed dilution by sequential streaking. These drops were allowed to equilibrate for 24hrs and only those conditions that produced predominantly clear drops were then streaked. Streaking was done using a cats whisker. This was first touched to a "best of the bad" apo crystal where it picked up minuscule fragments (nucleation seeds) and then successively pulled through the 3 'dilution' drops. The streaked drops were re-sealed over their reservoirs and the whisker rinsed in MilliQ water to remove mother liquor and excess seeds. The process was then repeated for the next condition.

2.3.16 Diffraction data collection

All diffraction testing and data collection were undertaken at 100 Kelvin (K) using a synchrotron source of X-rays, either at the UK Daresbury SRS, Warrington or Diamond Light Source, Didcot. Cryosolutions for crystal freezing were made using the same reservoir conditions that the crystal came from including, where appropriate, additives, co-enzyme and inhibitor. A range of glycerol concentrations within the cryosolution were tested to find the lowest concentration which produced a clear, glassy drop upon freezing. Using a crystal pin (Hampton Research or Spine) with a nylon loop of similar size to the crystal, a visually good crystal was selected and transferred from its drop into a drop of cryosolution. The crystal was then instantly re-looped and placed on the goniometer head in the path of a 100 K cryostream, ready for X-rays. Cryocooling crystals prior to data collection is known to reduce the problem of radiation damage and is of particular importance at a synchrotron source (reviewed in Garman and Owen, 2006). Test diffraction images were taken at various angles, exposure time and detector distance to determine; a) if data from the crystal was worth collecting, b) what combination of time and distance produced the 'best' resolution and c) was data at certain angles worse then others. Alternatively, the crystal was frozen on its pin by fast immersion in liquid nitrogen and transferred to a cold Dewar for storage or transport.

2.3.17 Data processing

The programs used for data processing in this thesis are incorporated into the HKL2000 suite (Otwinowski and Minor, 1997).

2.3.17.1 Data indexing and integration

The position of diffraction peaks are dictated by the unit cell and Bravais Lattice of the crystal and therefore their accurate identification is required for the determination of correct cell parameters. The program Denzo was used to automatically pick a small number (70 - 100) of spot positions. The accuracy of these picks was inspected by eye in the XDisplayF program, with some manual intervention for spot removal or addition where deemed necessary. Where multiple lattices were present efforts were made to only select stronger peaks, being from the major lattice. Rounds of refinement in

Denzo, initially in the primitive triclinic (P1) lattice, improves spot position leading to an improvement in cell parameters. This was done across a 3D window of 5 frames which was found to help with selection of the predominate lattice. Additionally, the areas designated to define the spot and the local background were visually inspected. Where necessary the size of these areas were altered to avoid including excess background in the spot region (spot size) and avoid nearby spots being included in the background definition (box size), maximising the signal to noise ratio. After refinement statistics stabilised, the Bravais Lattice table was inspected. Here, the spot positions are distorted to fit the parameters for each lattice type. The Bravais Lattice with the highest symmetry and requiring only minimal distortion was selected. Further rounds of refinement apply this lattice to the data and visual inspection is used to confirm spots are still being correctly picked.

A profile fitting radius was defined to include approximately 50 reflections across the 3D window. As the data is integrated the radius helps determine accurate diffraction intensities for each spot, as the profiles of each reflection are compared to other nearby spots.

2.3.17.2 Scaling

Integration of each reflection is done on an arbitrary scale, therefore scaling is performed to put the intensities of reflections with the same index on the same scale. This is important as effects such as the variable diffracting ability of the crystal in different orientations can alter the intensity of reflections across frames. Multiple rounds of scaling are implemented to refine the intensities (ScalePack). An error model is used to estimate the systematic errors in each resolution shell. This is adjusted using improved estimates, outputted as the linear *R*-factor in the ScalePack log file. The scale factor error value is also altered to produce an overall Chi^2 value near 1.0. Additionally, the data was defined to extend to the lowest resolution shell where the I/σ (intensity/error) is approximately 2. Data beyond this point was not used. Scaling results in a list of the positions (*hkl* indices) and intensities of reflections, giving a complete description of the diffraction data.

2.3.18 Molecular Replacement

Molecular replacement was accomplished using programs from the Collaborative Computational Project 4 suite (1994, CCP4). Data from ScalePack was first converted in CCP4 mtz format, using ScalePack2mtz. An mtz file contains *hkl* indices, intensity information, observed structure amplitudes (F_{obs}), a list of reflections and FreeR flags (for refinement analysis). To determine the number of protein monomers in the asymmetric unit (AU), the Matthews Cell Content Analysis program was used. This produces a Matthews coefficient and calculates the percentage solvent content of the asymmetric unit based on the monomeric molecular weight of the protein. The most likely number of monomers present can then be determined. Phaser version 1.3 (Storoni, McCoy and Read, 2004) was used to carry out the molecular replacement. Phaser works by taking a model of reasonable sequence homology (greater then 25%) and using maximum likelihood to determine its rotation and translation function. Probability statistics are used to select the most likely solution, where a "Z-score" (number of standard deviations above the mean value) of greater then 6 indicates a 'possible' solution. The solution is also refined by calculating and improving the Log Likelihood Gain (LLG), which measures how much better the data is predicted by the model as opposed to a random collection of the same atoms. This should increase as the model is improved and explains more of the data. Potential atom clashes within the solution can also be allowed for.

2.3.18.1 Molecular replacement of the ternary data

Only one protein structure was available with high enough sequence identity to be used in molecular replacement for the human GRHPR, the *H. methylovorum* D-glycerate dehydrogenase (Goldberg, Yoshida and Brick, 1994). This has a sequence identity of 34% and was available in the Protein Data Bank (PDB; http://www.rcsb.org) under the code 1GDH. This crystal structure had been solved in the apo form and contained a dimer in the asymmetric unit. Non-protein atoms (water and sulphate ions) were deleted from this file and searches were executed using either the dimer or one monomer chain as the search model. Further searches treated the two distinct domains known to be present in this family of enzymes - the substrate-binding domain, residues 1 to 99 and 291 to 321, and the coenzyme binding domain, residues 100 to 290 (numbering as in 1GDH) – as separate search models.

2.3.18.2 Molecular replacement of the apo data

The successful solution of the human GRHPR structure allowed this to be used as the molecular replacement model for the apo data. To minimise bias in the domain positions in the model, the substrate-binding domain (5 – 105, 300 – 328, numbering as in the human sequence) and the coenzyme binding domain (108 – 297) were used in separate sequential searches. Again, all non-protein atoms (water, sulphate ions and ligand atoms) were deleted from the ternary file prior to its use in molecular replacement.

2.3.19 Model building and refinement

Once a molecular replacement solution is found, theoretical structure factors can be calculated based on this model (F_{calc}). Electron density maps are calculated by Phaser (and later by Refmac5) based on the differences between the F_{obs} (Fo) and F_{calc} (Fc) and include phase information estimated from the model. Two maps are produced to assist in improving the model so that it explains the data better. The difference map (Fo-Fc) shows areas where atoms need to be added (positive density) and areas were model atoms need to be removed (negative density). As phase information is calculated from the model, density tends to be biased towards reproducing the model. A 2Fo-Fc map reduces this bias by strengthening the importance of the observed amplitudes and was used during model building. In these studies, visualisation of the maps and manual model alterations were performed in Coot (Emsley and Cowtan, 2004). Here the model was altered by the realignment of backbone atoms and the mutation and reorientation of side chains to better fit the density maps. During final stages of model building, ligand atoms, waters and other ions were also added where appropriate.

Interspaced with model building were rounds of refinement. This refines the model coordinate positions in an attempt to reduce the difference between the F_{obs} and F_{calc}. As the model now matches the data closer, improved density maps can be generated. These were fed back into model building for further improvements. The program Refmac5 (Murshudov, Vagin and Dodson, 1997) from the CCP4 suite was used for 10 cycles of restrained refinement after each stage of model building, outputting the refined model and both density maps. Where required, restraints were added to the

refinement process to ensure the root mean square (RMS) deviation of bond angles and lengths were retained at reasonable values. Furthermore, non-crystallographic restraints were used during early refinement to ensure that initially identical protein chains did not vary. Restraints were gradually loosened as the maps improved when the loosening had a positive effect on the R values. For a final round of refinement, the Translation/Libration/Screw (TLS) Motion Determination server (http://skuld.bmsc.washington.edu/~tlsmd; Painter and Merritt, 2006) was used to create a number of TLS groups for each monomer. This divides the protein chain into multiple fragments (groups) and models each as a rigid body undergoing vibrational motion. This produces a modified structure model file containing the TLS group definitions and a TLS tensor file. This is then refined in Refmac5. In addition, simulated annealing was attempted using the Crystallography and NMR System (CNS) software. However this proved ineffective in reducing the R values.

The success of refinement is shown by the convergence between the F_{obs} and F_{calc}, measured by the R_{cryst}. Additionally, an R_{free} is calculated from a subset (5%) of reflections not used during the refinement, providing an unbiased measure of model improvement. Once the R values, particularly the R_{free}, stopped decreasing the model was considered fully refined.

Unless otherwise stated all protein superpositions were undertaken using the function implemented in CCP4mg (Potterton *et al.*, 2002) and figures were made in PyMOL (DeLano, 2002).

2.3.20 Validation

During refinement the model was analysed by the inbuilt validation tools in Coot and later by SFCHECK (Vaguine, Richelle and Wodak, 1999) or MolProbity (using the all-hydrogen option, Davis *et al.*, 2007). Ramachandran outliners and bad rotamers were identified and improved. Any changes were followed by refinement of the new model.

2.3.21 Mutagenesis

Specific mutants were made in an effort to test the effect, via kinetic studies, of single amino acid changes on the substrate specificity of GRHPR. Mutations were created

using the QuickChange® Site-Directed Mutagenesis Kit (Stratagene). Primers were designed for each mutation; 29bp in length, starting and ending in a cytosine (C) or guanosine (G) and with the codon to be altered in the middle (Table 2.2). The mutant codon was selected so as to require the minimal change from the original.

Site	Primer	Sequence
L59	Original seq.	5'-GCCTGCTCTGCCTC**CTC**TCCGACCACGTG-3' …C L **L** S D…
	L59A_forward	5'-GCCTGCTCTGCCTC**GCC**TCCGACCACGTG-3' …C L **A** S D…
	L59V_forward	5'-GCCTGCTCTGCCTC**GTC**TCCGACCACGTG-3' …C L **V** S D…
S296	Original seq.	5'-CCCCACATTGGC**AGT**GCCACCCACAGAAC-3' …I G **S** A T…
	S296A_forward	5'-CCCCACATTGGC**GCT**GCCACCCACAGAAC-3' …I G **A** A T….

Table 2.2 Primers used in mutagenesis of GRHPR.

In all cases reverse primers are the complement of the forward.

The PCR reactions for each mutant were undertaken separately and consisted of 2.5U *PfuTurbo*® DNA Polymerase, 1x Reaction Buffer, 1µl dNTP mix (all supplied in the kit) and the forward and reverse primers (125ng of each). The pTrcHisB-T-GRHPR (50ng) was used as the template and the final reaction volume of 50µl was made up with Elgastat water. The PCR cycling protocol was implemented as in Table 2.3.

	Temperature	Time	No. of rounds
Initial DNA denaturation	95°C	30 seconds	1
Denaturation	95°C	30 seconds	16
Annealing	55°C	1 minute	
Amplification	68°C	5 minutes	
Hold	4°C	Indefinite	

Table 2.3 PCR protocol used during site directed mutagenesis of GRHPR.

Following PCR, 10U of the methylation specific restriction enzyme *Dpn*I was added to the reaction. This was incubated at 37°C for an hour, during which time the original template DNA is digested leaving only the mutated strands for transformation. A 1µl aliquot of the *Dpn*I digested mix was transformed into 50µl of XL1-Blue Supercompetent Cells (Stratagene) as described in section 2.3.5. Transformed cells

grew overnight at 37°C on tetracycline (10μg.ml^{-1})/ampicillin LB agar plates. A number of resulting colonies were selected, grown as 10ml liquid cultures and the plasmid purified. The successful incorporation of the mutations was confirmed by sequencing using the pTrcHis sequencing primers, pTrcHisF(orward) and R(everse) (Cogenics Inc). Additionally, the double mutant L59V/S296A was made by performing the same procedure using the S296A primers and an already confirmed L59V mutated plasmid as the template. Plasmids containing the correct mutation were then transformed into *E. coli* Rosetta, with expression and purification performed as for the native protein.

2.3.22 Kinetics

2.3.22.1 Native GRHPR

The enzyme activity was calculated by observing the change in absorbance at 340nm (Perkin Elmer spectrophotometer) due to the breakdown of the cofactor NADPH (or NADH) ($\varepsilon = 6220 M^{-1}.cm^{-1}$). All assays (triplicates) were run in 100mM potassium phosphate buffer pH 7.5, 37°C, with a final volume of 1mL. Substrates and cofactors were made up as required in the same buffer. The concentration of the cofactor (NADPH or NADH) was kept at 0.3mM in all assays and the substrate concentration was varied. Glyoxylate was varied from 0.02mM – 1mM, and hydroxypyruvate from 0.02mM - 0.5mM. Lineweaver-Burk plots ($1/V^{app}$ vs. 1/[substrate] as described by the equation below (Mathews, van Holde and Ahern, 2000), were used to determine the K_M and V_{max} of the reaction, where V^{app} is μmoles of NADH or NADPH consumed per minute.

$$\frac{1}{V^{app}} = \left(\frac{K_M}{V_{max}}\right)\left(\frac{1}{[S]}\right) + \frac{1}{V_{max}}$$

2.3.22.2 GRHPR Mutants

The kinetic studies of the mutant proteins were performed largely as for the native protein, except a 96-well UV spectrophotometer (Versamax, Molecular Devices) was used controlled by the SOFTmaxPro software. All assays were attempted with NADPH at 0.3mM in a final volume of 100μl. All potential substrates (glyoxylate, hydroxypyruvate and pyruvate) were varied from 100mM – 0.02mM.

Chapter 3 The structure of human GRHPR in 3 states

3.1 Chapter overview

This chapter describes the modification of the GRHPR expression vector, protein purification and subsequent crystallisation of the ternary, binary and apo states of this enzyme. Structure solution and refinement are described, followed by a discussion of the overall protein structure. Differences between the ternary and apo structures are discussed and comparisons made with the structures of related proteins. As the ternary structure of GRHPR represents the first structure of a D-2-hydroxyacid dehydrogenase family member to be solved as a true ternary complex, the model of substrate binding and stereospecificity for this family can be confirmed and is discussed. This chapter will include the proposition of a structure based model explaining the known substrate specificity of this enzyme, which is further discussed in Chapter 4. Finally, the chapter will conclude with an explanation of the potential effects of known PHII mutations. The structures of the human GRHPR ternary and binary states produced during this thesis have been published (Booth *et al.*, 2006) and is listed in Appendix 1.

3.2 Results

3.2.1 Molecular biology

3.2.1.1 T-GRHPR vector construction

The originally supplied GRHPR expression vector (pTrcHisB-HPR) resulted in the production of protein with a large N-terminal tag. This was likely to be detrimental for crystallisation, so PCR was used to insert the nucleotide sequence of the thrombin recognition site directly prior to the initial methionine of the GRHPR enzyme. The PCR product was electrophoresed through 1% agarose gels to separate the product from free dNTPs, primers and template. Figure 3.1a shows the result of the PCR, visualised due to the intercalation of the EtBr with the DNA producing UV fluorescence. After gel extraction and ligation into pGEM-Teasy, transformation into *E. coli* Nova Blue was carried out and the subsequent agar plates checked for white colonies. As expected, only 2 or 3 colonies were present on the negative control plate

and all were blue. This both indicates the absolute requirement of NovaBlue™ cells for ampicillin resistance provided by the vector and suggests that some pGEM-Teasy may be able to re-circularise, or is present in circular form in the commercial product. The agar plates from the PCR ligation showed approximately 50 colonies, of which less than 10 appeared blue. A number of white colonies were selected, recombinant plasmid DNA purified and digested to confirm the presence of the insert (Figure 3.1b). Vectors containing insert of the correct size were sequenced to establish that no mutations had been created during PCR. A sequenced insert was subsequently ligated into linearised pTrcHisB. The transformation of this into *E. coli* Rosetta produced a single colony. Again, recombinant plasmid DNA was prepared and digested confirming the creation of the pTrcHisB-T-GRHPR expression vector (Figure 3.1c).

3.2.2 Purification

The *E. coli* Rosetta strain contains the pRARE plasmid which provides the strain with extra tRNAs for codons rarely used in *E. coli* genes but more common in human genes. This makes it an excellent bacterial expression host for many human proteins. Small scale expression trials showed that the minor change to the vector and the change in host, compared to that already published (Rumsby and Cregeen, 1999), had had no detrimental effect and soluble protein was being produced (results not shown).

Initial purification of soluble T-GRHPR was carried out using standard Ni^{2+} affinity chromatography. Once bound to the affinity column and non-bound protein had been washed off, a shallow gradient of imidazole (20mM – 200mM) was used to elute the recombinant protein and separate it from any non-specific/weakly bound proteins. The elution profile of the protein was tracked by A_{280} measurements. The principal peak elutes between imidazole concentrations of 150mM and 200mM and is preceded by a considerably smaller shoulder peak (Figure 3.2). Analysis by SDS-PAGE indicates that the major peak is comprised predominantly of recombinant GRHPR, which has a monomeric molecular weight of 40kDa (Figure 3.3). The shoulder peak contains proteins non-specifically bound to the affinity column but the GRHPR peak is largely pure with some contaminating proteins present at around 24kDa. Fractions containing this largely pure T-GRHPR were pooled and carried through to the next purification step.

Thrombin digestion was carried out overnight on those fractions deemed of acceptable purity. Scrutiny by SDS-PAGE of a cleaved aliquot showed no uncleaved recombinant protein (not shown). Purification proceeded via Ni^{2+} affinity and benzamidine columns, connected in series. The Ni^{2+} column further purified the sample by removing any remaining (if negligible) uncleaved T-GRHPR and some of the non-specific proteins while the benzamidine column removed the added thrombin (Figure 3.4).

Concentration of the resulting flow through and buffer exchanging prepared the now tag-less GRHPR for final purification by Superdex 75 gel filtration chromatography. By concentrating the sample to a small volume (0.5ml) and running the column slowly ($0.3ml.min^{-1}$) a high degree of resolution could be achieved. The UV absorbance (A_{280}) trace shows a large, symmetrical peak which corresponds to an elution volume of 56.5ml (Figure 3.5) and purified GRHPR (Figure 3.6). By previously running a number of know molecular weight standards down the column, an equation had been produced correlating the elution volume to the log of the molecular weight of the protein. This equation is given below and is used for all further volume to molecular weight conversions for the Superdex 75 gel filtration column:

$$\log Mr = -1.4758(V_{elution}/V_{viod}) + 6.8879, \text{ where } V_{viod} = 41$$

Consequently the elution volume corresponds to a molecular weight of approximately 71kDa and confirms the dimeric nature of GRHPR (predicted 72kDa) in solution. Inspection of the GRHPR on SDS-PAGE indicates a greater the 95% purity, making it suitable for crystallisation trials. Purified protein could be stored at -20°C for long periods if needed, but was generally used 'fresh'.

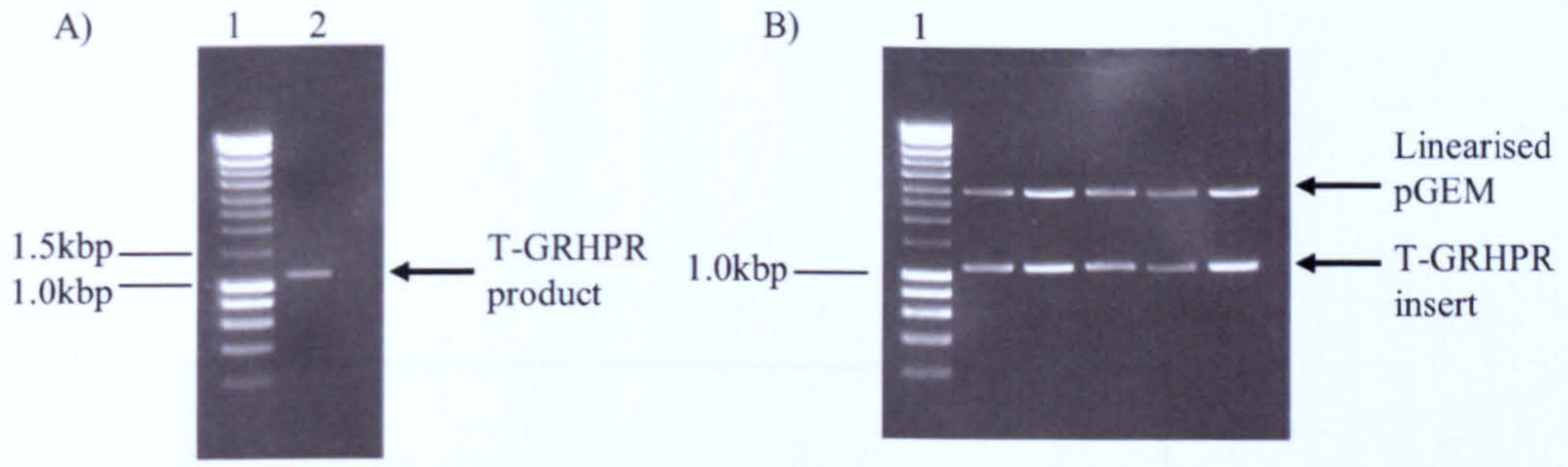


Figure 3.1 Agarose gels of T-GRHPR vector construction.

1% agarose gels of; Lane 1) DNA ladder with relevant sizes indicated, A) PCR product T-GRHPR (2), B) pGEM-Teasy digestions from selected white NovaBlue™ cells. The integrity of these inserts was confirmed by DNA sequencing, with one being selected to be cloned into pTrcHisB (4.4kbp), C) double digestion of final pTrcHisB-T-GRHPR vector showing presence of insert (2). All show the correct sized insert of approximately 1.1kbp.

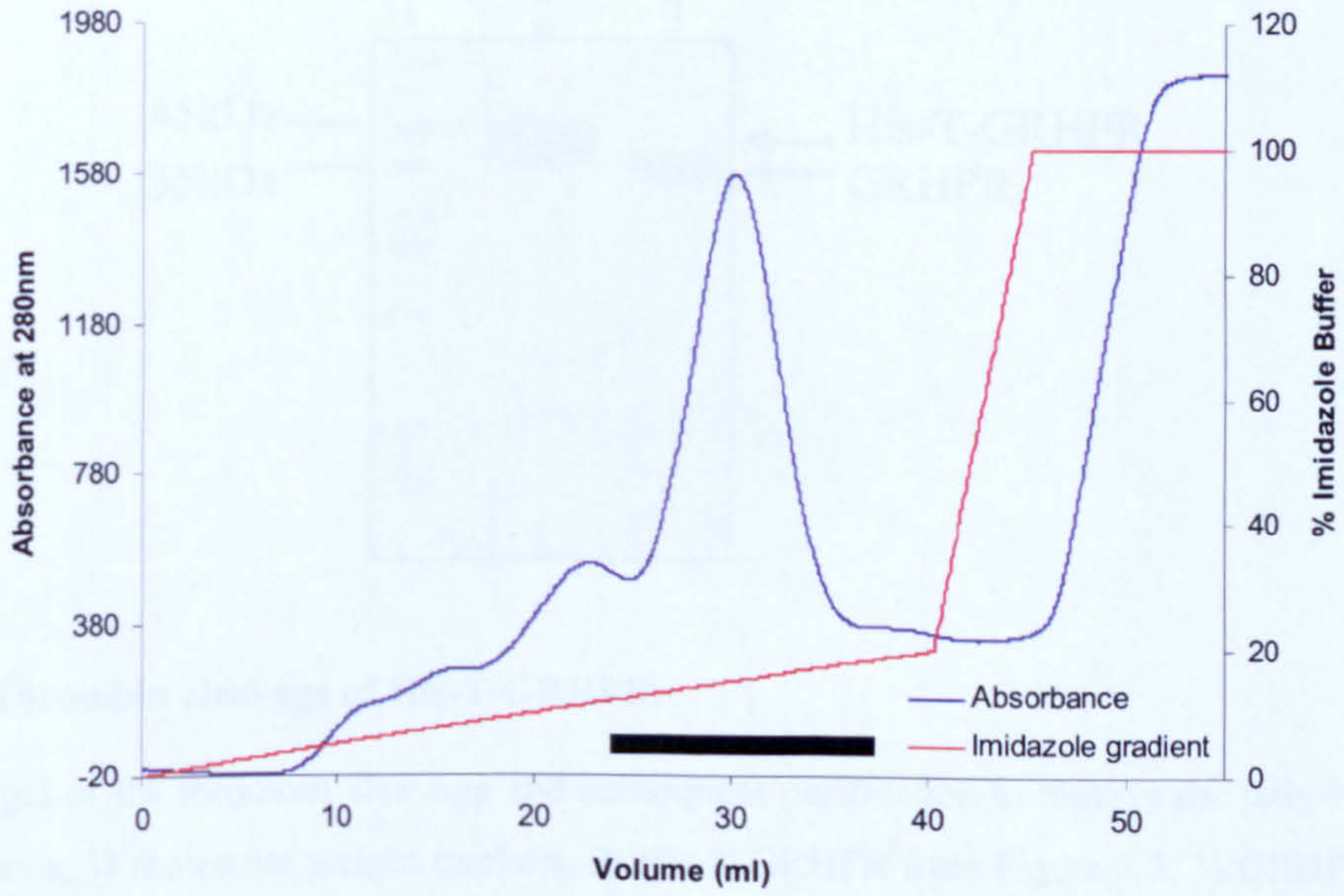


Figure 3.2 Nickel affinity chromatography trace of hexa-His tagged T-GRHPR.

The elution of protein was detected by absorbance at 280nm and the solid line corresponds to fractions shown in Figure 3.3. The imidazole gradient used for elution is also shown.

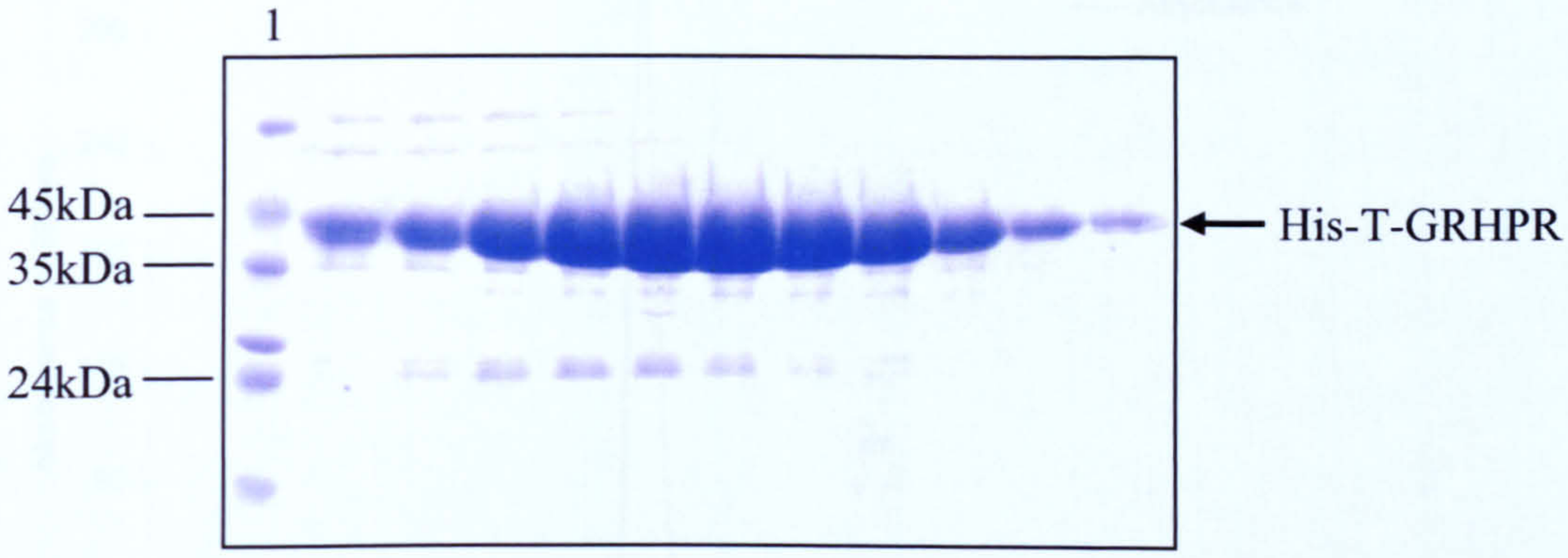


Figure 3.3 SDS-PAGE of the nickel affinity purification of His-T-GRHPR.

First step purification of the thrombin cleavable, poly-His tagged GRHPR. Lane 1 contains molecular weight markers with all other lanes corresponding to the region indicated by the solid line in Figure 3.2.

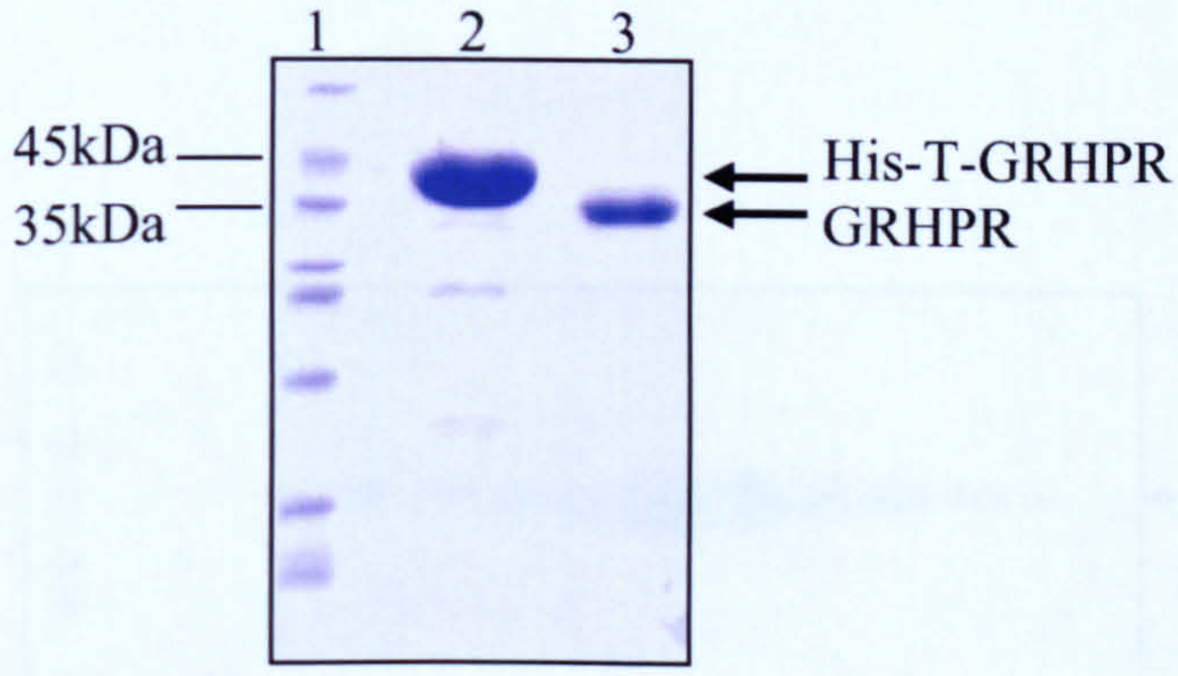


Figure 3.4 Thrombin cleavage of His-T-GRHPR.

SDS-PAGE gel of the thrombin cleavage and subsequent purification to remove the poly-His tag from GRHPR. Lanes; 1) molecular weight markers, 2) His-T-GRHPR from Figure 3.3, 3) GRHPR after His-tag removal and benzamidine column purification.

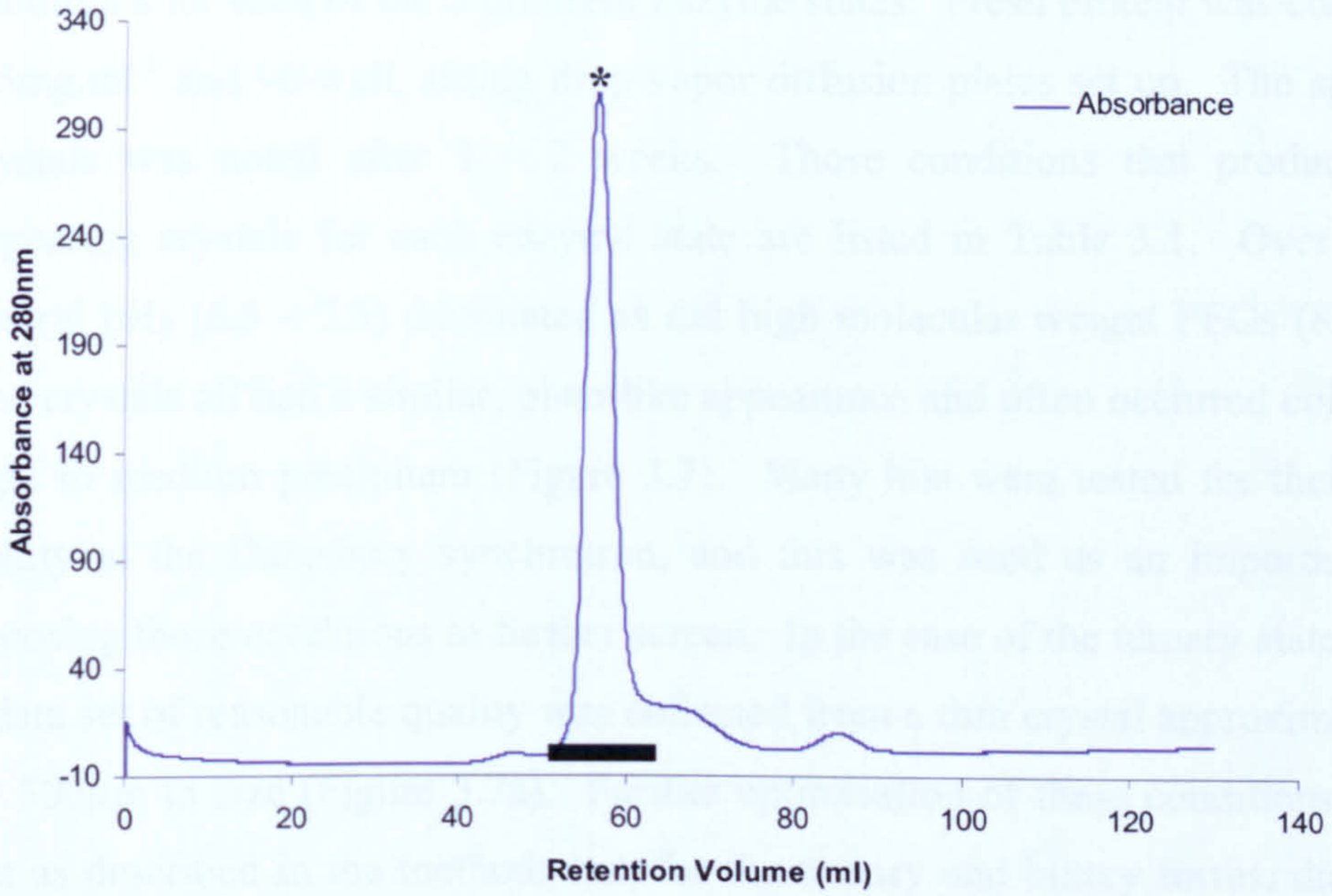


Figure 3.5 Size exclusion chromatography trace of human GRHPR.

Cleaved protein from Figure 3.4 was concentrated and further purified via size exclusion chromatography. The elution of protein was detected by absorbance at 280nm and the solid line corresponds to fractions shown in Figure 3.6. The star indicates the peak point corresponding to a retention volume of 56.5ml.

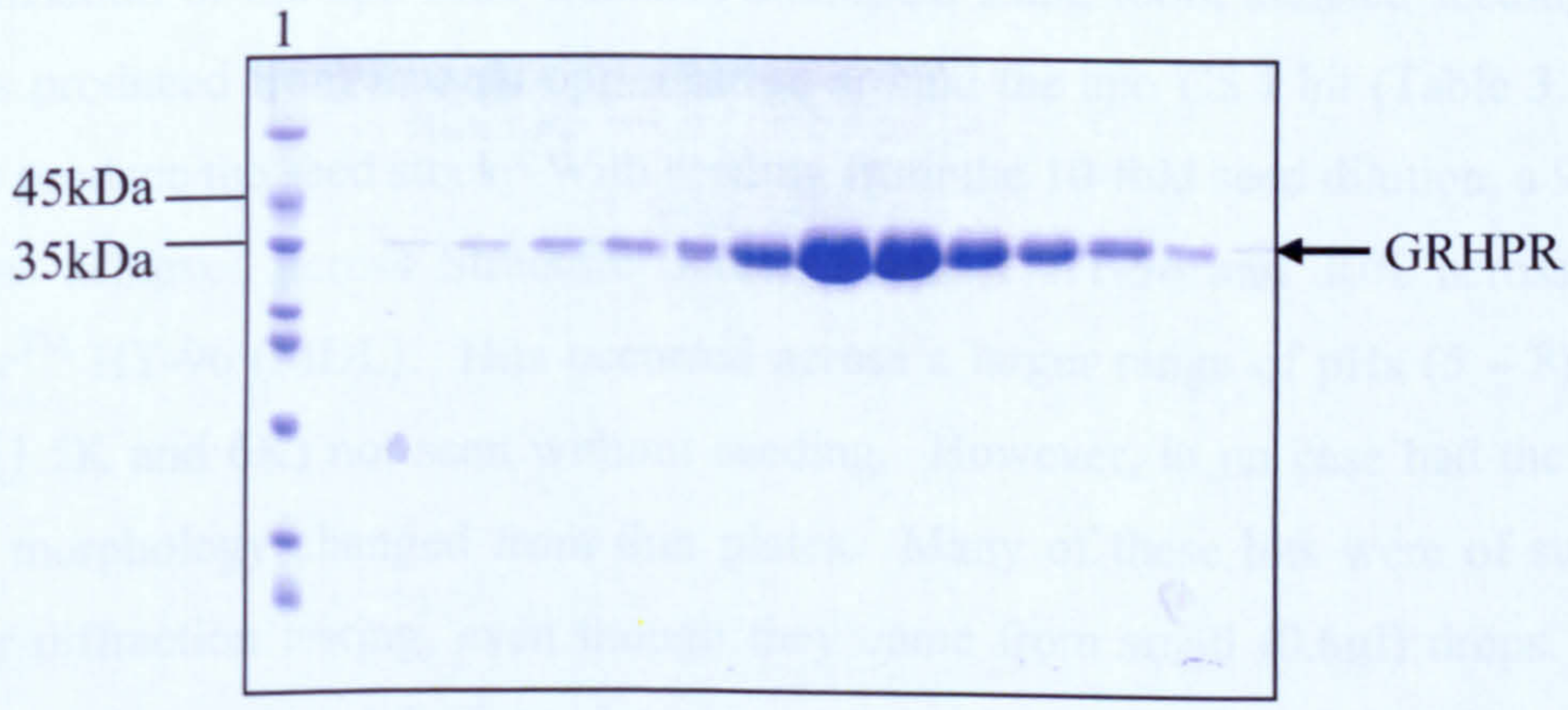


Figure 3.6 SDS-PAGE of size exclusion chromatography of GRHPR.

Final purification of human GRHPR; lane 1) molecular weight markers. Other lanes correspond to fractions marked with a solid line in Figure 3.5.

3.2.3 Crystallisation

Extensive crystallisation screening was carried out in an attempt to find the best conditions for each of the 3 different enzyme states. Fresh protein was concentrated to 5.5mg.ml^{-1} and 96-well, sitting drop vapor diffusion plates set up. The appearance of crystals was noted after 1 – 2 weeks. Those conditions that produced the best appearing crystals for each enzyme state are listed in Table 3.1. Over all the hits, neutral pHs (6.5 – 7.5) dominated as did high molecular weight PEGs (8K and 20K). The crystals all had a similar, plate-like appearance and often occurred coincident with light to medium precipitant (Figure 3.7). Many hits were tested for their diffraction ability at the Daresbury synchrotron, and this was used as an important guide for choosing those conditions to further screen. In the case of the ternary state CS Lite hit, a data set of reasonable quality was collected from a thin crystal approximately 200µm by 500µm in size (Figure 3.7a). Further optimisation of these conditions was carried out as described in the methods but, for the ternary and binary forms, did not further improve crystal quality. Additionally, screens carried out with the higher concentrations of NADPH and oxalate (2mM and 5mM respectively) only produced very poor quality hits. As a comparison between the different states, it was noted that within the 96 conditions of Crystal Screen I and II (Hampton Research) the ternary complex produced crystallisation hits in 10.4% of conditions, the binary complex in 8.3% and the apo form in 5.2%.

Crystallisation of the apo state was also attempted using robot assisted seeding. Apo crystals produced from manual optimisation around the apo CS I hit (Table 3.1) were used to produce the seed stock. With seeding from the 10-fold seed dilution, a 9.4% hit rate was achieved across Structure Screen I and II HT-96 and 30% across PACT *premier*™ HT-96 (MDL). Hits occurred across a larger range of pHs (5 – 8) and in PEGs (1.5K and 6K) not seen without seeding. However, in no case had the overall crystal morphology changed from thin plates. Many of these hits were of sufficient size for diffraction testing, even though they came from small (0.6µl) drops. The 2 conditions that produced the best crystals upon seeding are given in Table 3.2. Crystals from these conditions diffracted to a low resolution, however they could not be reproduced when manual optimisation was attempted. The best improvement of apo crystals came with lowering the protein concentration and streaking. This was largely concentrated around the CS I condition hit (Table 3.1). Initial fine pH, PEG and

magnesium screening resulted in reproducible crystals in 0.1M sodium cacodylate pH 6.75, 16% PEG 8K and 0.25M magnesium acetate tetrahyrate, which diffracted to 3.5Å – 4Å. Using this starting point and these crystals as the streak source, crystals were still produced even when the protein concentration had been lowered to 1mg.ml^{-1}. Further improvements were made by lowering the PEG percentage in the streaked drops. In contrast, lowering the magnesium concentration resulted in heavy precipitant. The best crystal was eventually produced by streaking into drops setup with 3.5mg.ml^{-1} of GRHPR and 0.1M sodium cacodylate pH 6.75, 10% PEG 8K and 0.25M magnesium acetate tetrahyrate as the mother liquor (Figure 3.7).

Form	Screen	Buffer/pH	Precipitant	Salt/Additive
Ternary	CS Lite	0.1M sodium cacodylate pH 6.5	15% PEG 8K	0.2M ammonium sulphate
	JBS 1	0.1M HEPES pH 7.5	25% PEG 1K	none
	CS II	0.1M MES pH 6.5	12% PEG 20K	none
Binary	CS I	0.1M sodium cacodylate pH 6.5	18% PEG 8K	0.2M calcium acetate hydrate
	CS II	none	1.6M tri-sodium citrate dehydrate pH 6.5	none
	JBS 5	0.1M MES pH 6.5	10% PEG 20K	none
Apo	CS I	0.1M sodium cacodylate pH 6.5	20% PEG 8K	0.2M magnesium acetate tetrahyrate
	JBS 1	0.1M MES pH 6.5	15% PEG 400	none
	CS II	0.1M HEPES pH 7.5	1.6M ammonium sulphate	0.1M sodium chloride

Table 3.1 Top 3 hits for each crystal form.

The ternary and binary hits shown here are the result of crystal screens set up with the lower concentrations of NADPH and oxalate (Table 2.1). All protein concentrations were 5.5mg.ml^{-1}.

Screen	Condition
PACT *premier*™	0.1M Bis-Tris-HCl propane pH 6.5, 0.2M sodium bromide, 20% w/v PEG 3350
PACT *premier*™	0.1M Succinic acid, Phosphate, Glycine system pH 6.0, 25% w/v PEG 1500

Table 3.2 Top apo crystallisation conditions identified during seeding.

Both conditions were identified during a 10-fold seed dilution screen.

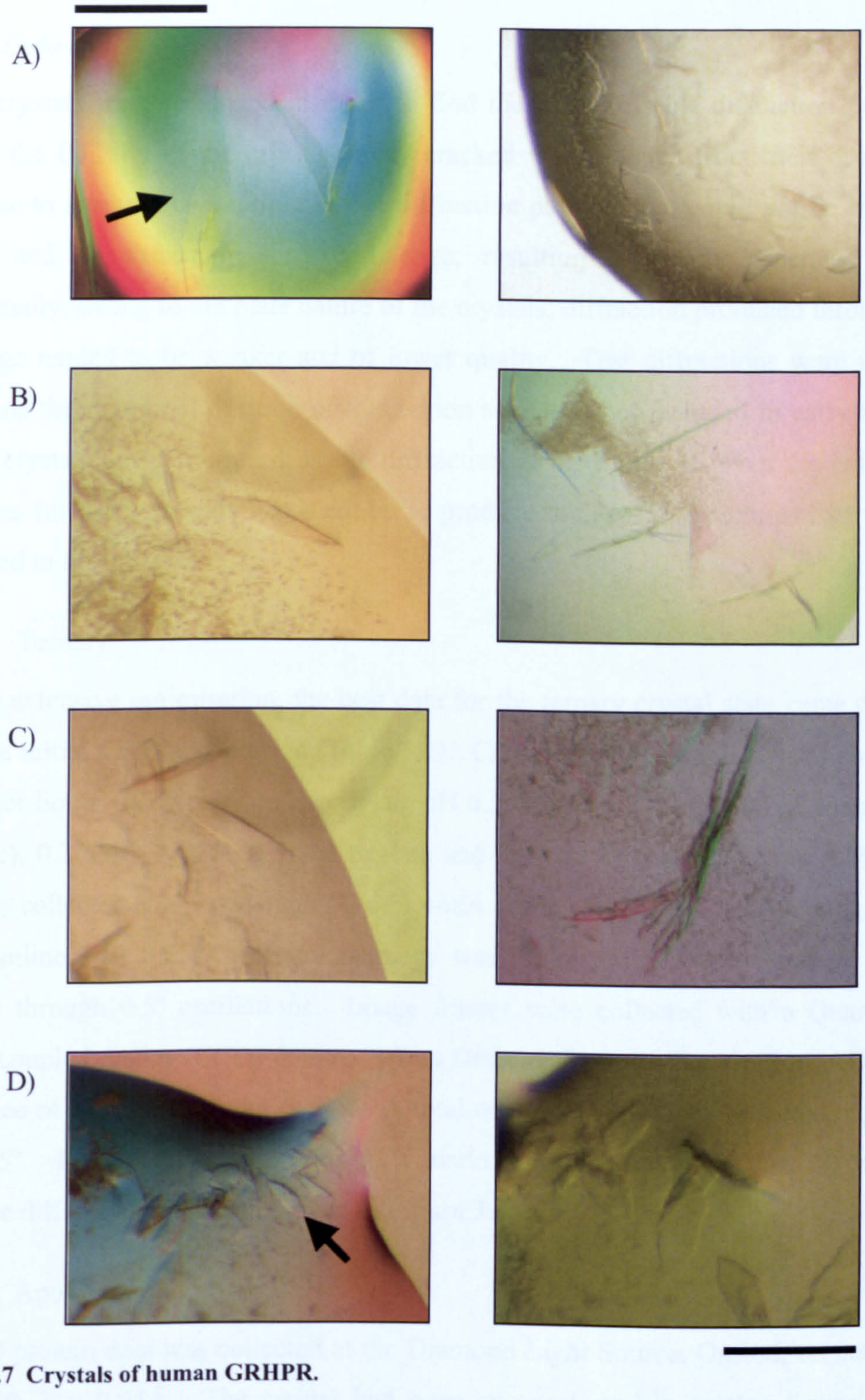


Figure 3.7 Crystals of human GRHPR.

Photographs of crystals relating to the top two conditions (left and right columns) given in Table 3.1; A) ternary, B) binary, C) apo. D) shows GRHPR apo crystals along the path of a streak. Final data came from the crystals indicated by arrows: A) Ternary; 5.5mg.ml^{-1}, 0.1M sodium cacodylate pH 6.5, 15% PEG 8K and 0.2M ammonium sulphate, D) Apo; streaking into 3.5mg.ml^{-1}, 0.1M sodium cacodylate pH 6.75, 10% PEG 8K and 0.25M magnesium acetate tetrahyrate. The scale bar in the top left represents 200μm and is applicable only to that crystal. The bar in the lower right represents 100μm and is applicable to the remaining photographs.

3.2.4 Data collection

Many crystals were irradiated in order to find those with usable diffraction patterns. Due to the fragility of the crystals many cracked when looped from their drops and gave rise to smeared or multiple lattice diffraction patterns that could not be indexed. Others had an extremely long cell edge, resulting in poorly separated spots. Additionally, owing to the plate nature of the crystals, diffraction produced through the thin edge tended to be weaker and of lower quality. Test diffractions were used to determine the rotational position of this region so it was not included in early frames. Binary crystals never produced usable diffraction data. Below is given the collection strategies for those crystals that went on to produce the protein structures that will be discussed in this chapter.

3.2.4.1 Ternary

Despite extensive optimisation, the best data for the ternary crystal state came directly from the initial CS Lite condition (Table 3.1). Cryoprotectant for this crystal consisted of mother liquor (0.1M sodium cacodylate pH 6.5, 15% PEG 8K and 0.2M ammonium sulphate), 0.2mM NADPH, 0.5mM oxalate and 30% w/v glycerol. X-ray diffraction data was collected at a wavelength (λ) of 1.448Å at the Daresbury SRS, Warrington, on PX beamline 14.1. Each diffraction image was the result of X-ray exposure for 20 seconds through 0.5° oscillations. Image frames were collected with a Quantum 4 charge-coupled device (CCD) detector (Area Detector Systems Corporation, ADSC) at a distance of 120mm from the crystal. A total of 319 images were collected, equating to 159.5°. Protein diffraction showed a marked decrease over the last 60 images. Example diffraction images are given in Figure 3.8.

3.2.4.2 Apo

The apo protein data was collected at the Diamond Light Source, Oxford, on beamline IO3, with λ = 0.95Å. The crystal had been cryoprotected in mother liquor (0.1M sodium cacodylate pH 6.75, 10% PEG 8K and 0.25M magnesium acetate tetrahyrate) with 30% w/v glycerol. Exposure to X-rays for 3 seconds through 1° oscillations produced usable data. A total of 360 diffraction images were collected with a Quantum 315, a 3x3 CCD detector (ADSC) at a distance of 300mm. An example diffraction image is given in Figure 3.9.

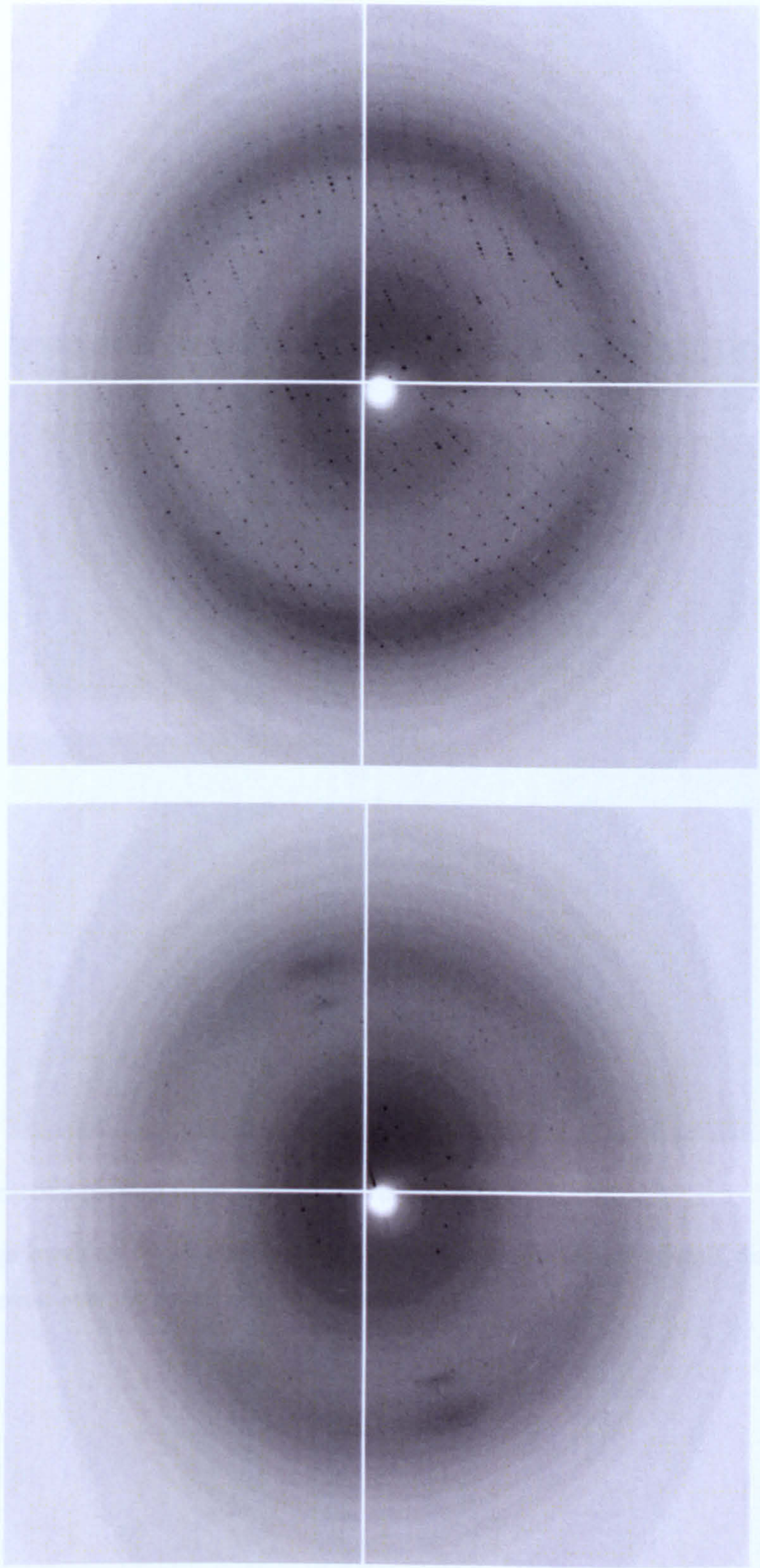


Figure 3.8 Diffraction pattern from ternary GRHPR crystals.

The upper image shows the data quality early in the collection, with the long cell edge evident by the closely spaced spots. The lower image shows the marked decrease in diffraction after 260 images (130°).

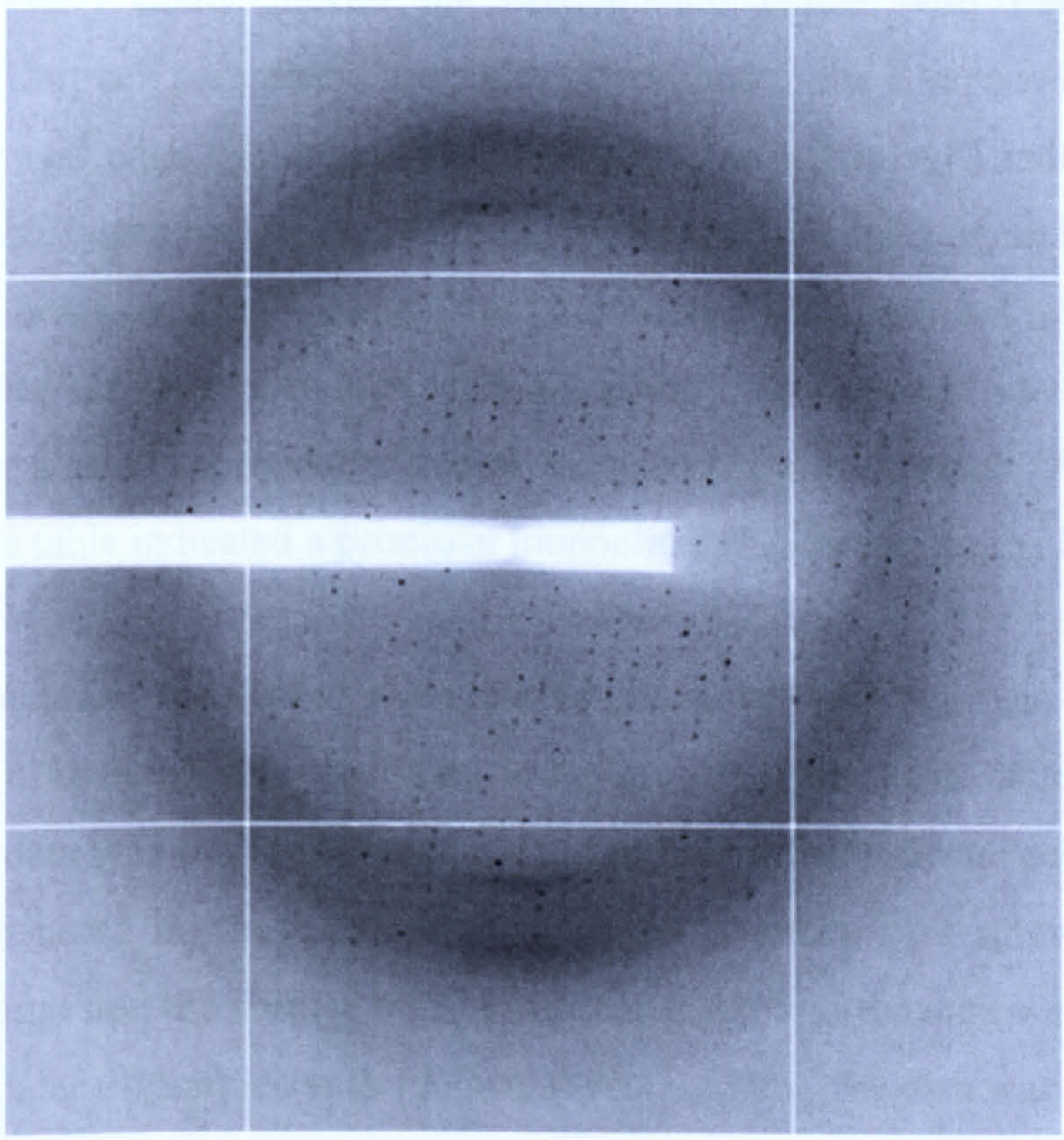


Figure 3.9 Diffraction pattern from the apo GRHPR crystal.

The long cell edge is evident by the closely spaced spots. Unlike the ternary crystals, diffraction did not show a large decrease over the course of the collection.

3.2.5 Data processing

3.2.5.1 Ternary data

Despite the presence of what appeared to be a weakly diffracting secondary lattice, the automatic indexing procedure within Denzo (part of the HKL2000 suite, Otwinowski and Minor, 1997) easily picked the predominate lattice with minimal manual intervention. Unit cell parameters were initially refined in the P1 space group with the expected long cell edge apparent. Visual scrutiny (in XDisplayF) indicated that the default "box size" used by Denzo was too large for the spot spacing, and was resulting in many spots being deemed 'overlaps' and discounted. A significant reduction in box size and further refinement rounds lead to a large improvement in the refinement statistics, more accurate spot identification and a decrease in overlaps. Analysis of the Bravais Lattice table indicated a primitive monoclinic (P2) space group (Table 3.3).

Subsequent rounds of refinement were carried out in the P2 lattice type to the edge of the detector. Predicted spot positions were seen to match the diffraction pattern closely. The data was then integrated in P2 and scaled using ScalePack. Systematic absences were noted on the *k* axis, indicating the presence of a screw axis (Figure 3.10). This suggested the correct space group to be $P2_1$ and the data was reindexed as such. Due to later difficulties with molecular replacement, the data was reindexed and scaled using only the first 250 frames, removing those frames that showed a significant loss of diffraction. This resulted in improved scaling statistics and data that enabled the structure to be successfully solved. A final resolution of 2.2Å was achieved, giving a completeness of 90.3% and an R_{sym} of 12.1%. All final scaling statistics are given in Table 3.5.

Lattice	Metric tensor distortion index	Best cell (symmeTris-HCled) Best cell (without symmetry restrains)					
primitive cubic	21.81%	66.72	77.90	151.26	79.97	91.45	91.12
		98.62	98.62	98.62	90.00	90.00	90.00
I centred cubic	31.34%	151.26	101.56	170.25	89.40	143.02	96.72
		141.03	141.03	141.03	90.00	90.00	90.00
F centred cubic	31.87%	191.80	170.25	172.05	45.88	112.38	129.45
		178.03	178.03	178.03	90.00	90.00	90.00
primitive rhombohedral	14.03%	151.26	157.62	163.76	36.78	24.03	29.12
		157.55	157.55	157.55	29.98	29.98	29.98
		72.31	72.31	450.32	90.00	90.00	120.00
primitive hexagonal	14.31%	66.72	77.90	151.26	79.97	91.45	91.12
		72.31	72.31	151.26	90.00	90.00	120.00
primitive tetragonal	5.48%	77.90	66.72	151.26	88.55	79.97	88.88
		72.31	72.31	151.26	90.00	90.00	90.00
I centred tetragonal	6.49%	77.90	66.72	304.97	101.46	94.99	88.88
		72.31	72.31	304.97	90.00	90.00	90.00
primitive orthorhombic	4.22%	66.72	77.90	151.26	79.97	91.45	91.12
		66.72	77.90	151.26	90.00	90.00	90.00
C centred orthorhombic	2.12%	77.90	298.96	66.72	88.83	91.12	85.16
		77.90	298.96	66.72	90.00	90.00	90.00
I centred orthorhombic	5.20%	66.72	77.90	304.97	94.99	78.54	91.12
		66.72	77.90	304.97	90.00	90.00	90.00
F centred orthorhombic	6.50%	101.56	103.55	304.97	101.15	86.34	81.16
		101.56	103.55	304.97	90.00	90.00	90.00
primitive monoclinic	**0.70%**	**77.90**	**66.72**	**151.26**	**91.45**	**100.03**	**88.88**
		77.90	**66.72**	**151.26**	**90.00**	**100.03**	**90.00**
C centred monoclinic	2.06%	298.96	77.90	66.72	91.12	91.17	94.84
		298.96	77.90	66.72	90.00	91.17	90.00
primitive triclinic	0.00%	66.72	77.90	151.26	79.97	88.55	88.88

Table 3.3 Bravais lattice table for the ternary GRHPR data.

Output from *DENZO* (implemented within the *HKL2000* data processing suite of programs) showing the most probable lattice type (bold), from the complete set of possible Bravais lattices.

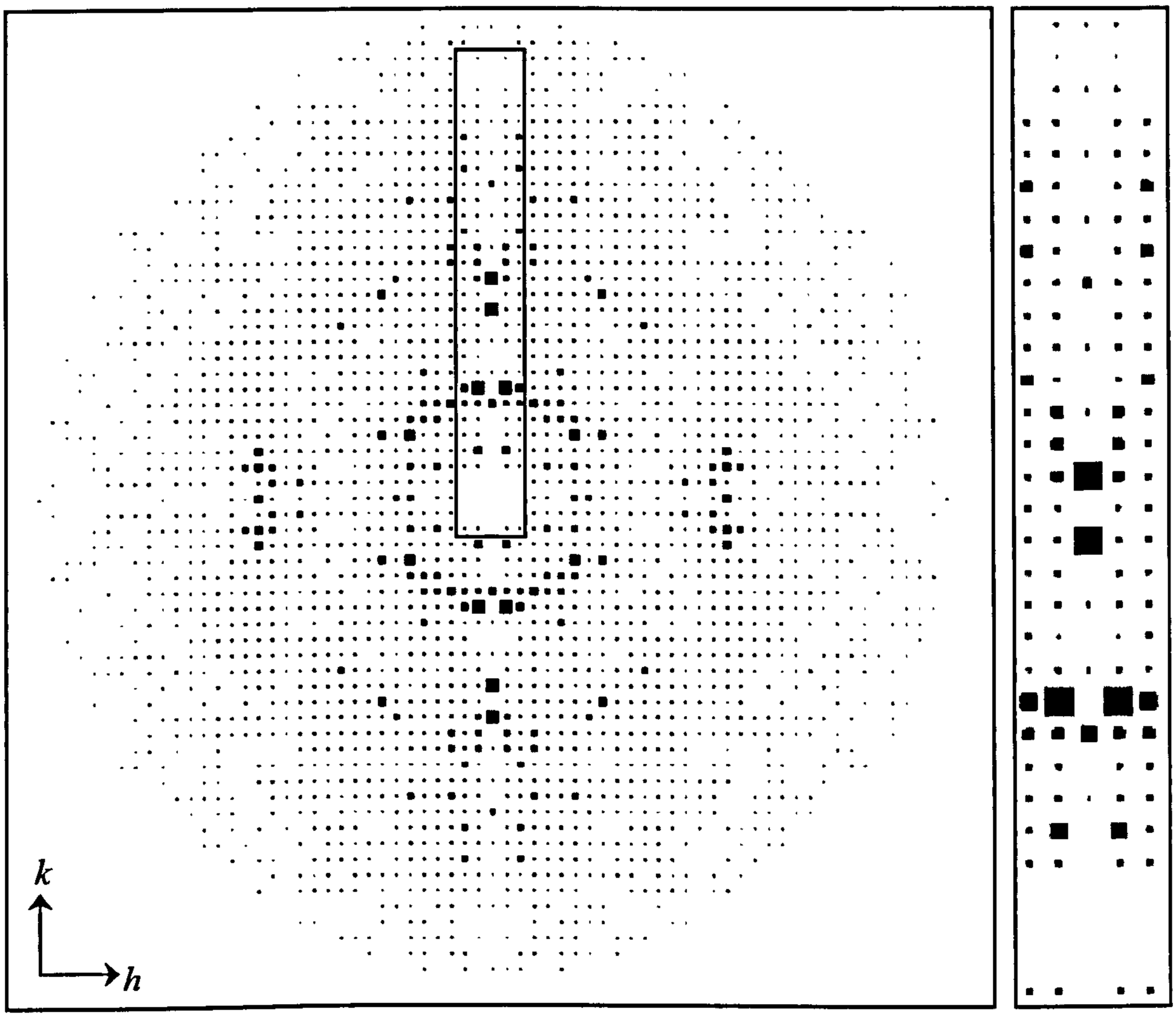


Figure 3.10 Pseudo-precession image of the ternary data processed in space group P2.

Image from the plane $l = 0$. The vertical image (right hand side) shows a close-up view of the intensities present along the *k*-axis and corresponds to the smaller boxed region. Intensities are visible every second position only, indicating the presence of a screw-axis along *k* and thus diagnosing the space group to be $P2_1$. The image was produced using HKLVIEW (CCP4 suite; 1994, CCP4).

3.2.5.2 Apo data

Unlike the ternary data, the automatic indexing procedure within Denzo failed to consistently pick a reasonable lattice and required some manual intervention. By manually removing a number of spot predictions clearly from a secondary lattice and the selection of clear lines of others, a reasonable lattice was produced. This went on to be successfully refined in P1, producing a unit cell clearly different to that of the ternary form. Again, due to the long cell edges the spots were very close together and the default box size required adjustment. Initial refinement and inspection of the Bravais Lattice table indicated P2 as the best lattice type. Further refinement, now in this lattice type, indicated the higher symmetry primitive orthorhombic (P222) lattice was appropriate (Table 3.4). Subsequent refinement in P222 produced well picked spots. The data was then integrated and scaled using ScalePack. The P222 lattice has a number of possible space groups (P222, $P222_1$, $P2_12_12$ and $P2_12_12_1$) due to the possibility of various screw axes. Initial scaling was carried out in P222 and systematic absences investigated. Absences were clearly visible on the *l* and *k* axes and potentially the *h* making the likely space group was $P2_12_12_1$. However, due to some violations of the expected absences $P2_12_12$ could not be ruled out (Figure 3.11). Incorrect space group identification can have severe ramifications on the ability to find a molecular replacement solution. Therefore, the data was scaled as both $P2_12_12$ and $P2_12_12_1$ and molecular replacement undertaken. A solution was successfully found only in the $P2_12_12_1$ space group, indicating this to be correct. A final resolution of 2.85Å was used, giving data completeness of 91.0% and an R_{sym} of 18.4%. All final scaling statistics are given in Table 3.5.

Lattice	Metric tensor distortion index	Best cell (symmeTris-HCled) Best cell (without symmetry restrains)
primitive cubic	27.41%	195.41 66.59 114.46 90.00 90.43 90.00 125.49 125.49 125.49 90.00 90.00 90.00
I centred cubic	40.30%	206.45 225.73 132.42 115.64 98.98 35.33 188.20 188.20 188.20 90.00 90.00 90.00
F centred cubic	41.21%	235.35 236.76 235.35 58.01 147.13 111.75 235.82 235.82 235.82 90.00 90.00 90.00
primitive rhombohedral	20.96%	206.45 225.73 195.41 30.47 18.82 35.33 209.20 209.20 209.20 28.20 28.20 28.20 99.50 99.50 600.18 90.00 90.00 120.00
primitive hexagonal	18.95%	195.41 114.46 66.59 90.00 90.00 90.43 154.94 154.94 66.59 90.00 90.00 120.00
primitive tetragonal	12.88%	195.41 114.46 66.59 90.00 90.00 89.57 154.94 154.94 66.59 90.00 90.00 90.00
I centred tetragonal	15.17%	114.46 66.59 411.84 99.30 105.71 90.00 90.53 90.53 411.84 90.00 90.00 90.00
primitive orthorhombic	**0.18%**	**66.59 114.46 195.41 89.57 90.00 90.00** **66.59 114.46 195.41 90.00 90.00 90.00**
C centred orthorhombic	4.02%	66.59 396.46 114.46 89.58 90.00 99.67 66.59 396.46 114.46 90.00 90.00 90.00
I centred orthorhombic	7.83%	66.59 114.46 411.84 105.71 80.70 90.00 66.59 114.46 411.84 90.00 90.00 90.00
F centred orthorhombic	7.93%	66.59 238.42 396.46 86.91 80.33 73.78 66.59 238.42 396.46 90.00 90.00 90.00
primitive monoclinic	0.00%	114.46 66.59 195.41 90.00 90.43 90.00 114.46 66.59 195.41 90.00 90.43 90.00
C centred monoclinic	4.02%	396.46 66.59 114.46 90.00 90.42 80.33 396.46 66.59 114.46 90.00 90.42 90.00
primitive triclinic	0.00%	66.59 114.46 195.41 89.57 90.00 90.00

Table 3.4 Bravais lattice table for the apo GRHPR data.

Output from *DENZO* (implemented within the *HKL2000* data processing suite of programs) showing the most probable lattice type (bold), from the complete set of possible Bravais lattices.

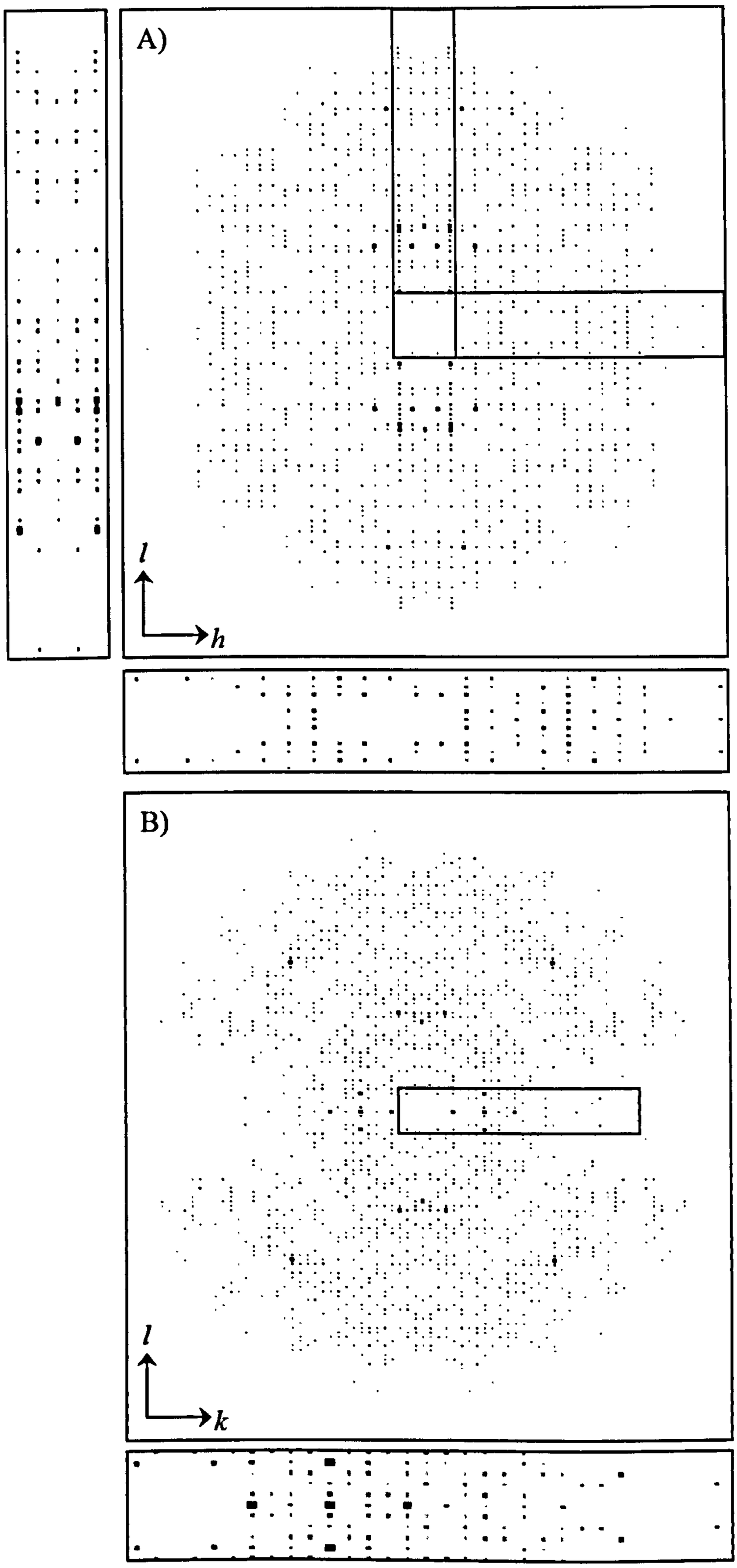


Figure 3.11 Pseudo-precession image of the apo data processed in space group P222.

(Legend on proceeding page)

Figure 3.8 Upper image (A) from the plane $k = 0$. The vertical image (left hand side) shows a close-up view of the intensities present along the *l*-axis while the horizontal image (middle) shows a close-up view of the intensities present along the *h*-axis. Both correspond to the smaller boxed regions in (A). Lower image (B) from the plane $h = 0$. The horizontal image (bottom) shows a close-up view of the intensities present along the *k*-axis and corresponds to the smaller boxed region of (B). In all cases intensities are visible every second position only, indicating the likely presence of three screw-axes and the space group to be $P2_12_12_1$. The image was produced using HKLVIEW (CCP4 suite; 1994, CCP4).

	Ternary	**Apo**
Space group	$P2_1$	$P2_12_12_1$
Unit cell (Å)	a = 76.6, b = 66.9, c = 149.8 β = 98.22	a = 67.36, b = 116.08, c = 198.23
Resolution range (Å)	50 - 2.2 (2.28 – 2.2)	50 - 2.85 (2.95– 2.85)
Number of unique reflections	69056 (5483)	34609 (3272)
I/σ	8.4 (2.1)	10.5 (2.0)
R_{sym} (%)	12.1 (27.5)	18.4 (79.1)
Completeness (%)	90.3 (72.3)	91.0 (87.6)
Redundancy	2.6 (2.2)	11.1 (8.8)
Mosaicity (°)	0.586	1.11

Table 3.5 Final scaling statistics for both the ternary and apo data sets.

Values are for the complete dataset; bracketed values represent the highest resolution shell only.

3.2.6 Molecular Replacement

3.2.6.1 Ternary

The ScalePack output file containing the *hkl* indices and intensities of reflections was converted into the CCP4 mtz format and the contents of the cell analysed, shown in Table 3.6. Considering the dimeric nature of GRHPR, a Matthews coefficient of 2.7 and a solvent content of 53.6%, the asymmetric unit was predicted to contain 4 monomers (2 dimers).

Finding a molecular replacement solution for the ternary data proved difficult. Initial searches using either the *H. methylovorum* D-glycerate dehydrogenase (GRHPR) dimer to search for 2 copies or using the monomer searching for 4 copies failed. Large rigid domain movements are known to occur within this family of enzyme between the apo and ternary conformations. Therefore, despite the reasonable sequence identity, the

apo structure could not automatically be considered a good model for the ternary data. To compensate for these movements the model structure was split into its component domains and successive rounds of Phaser searches and refinement were used to find a solution (Figure 3.12). The two coenzyme binding domains (100 - 290) make up over 60% of the dimer and were retained as a single search model. This domain shows 35% sequence identity with that of the human GRHPR and Phaser was used to search for 2 copies of it within the asymmetric unit. This produced a weak solution for the first copy but improved upon addition of the second (Phaser round 1, Figure 3.12 and Table 3.7a). Investigation of the resulting maps in Coot indicated reasonably continuous density for the backbone of the model. Some model building and refinement, including mutating the sequence to that of the human protein, was carried out and is described in detail in section 3.2.7.1. A subsequent Phaser search used a refined human coenzyme-binding domain dimer and a poly-alanine model of the bacterial substrate-binding domain (1 to 99, 291 to 321). This successfully found good solutions for two of the expected four substrate-binding domains (Phaser round 2, Figure 3.12 and Table 3.7b). The Z-scores of 7.0 and over and the progressive increase in LLG indicate the success of this. Supporting this, Phaser had positioned the breaks of the domains near each other and they were linked by continuous Fo-Fc density. The missing amino acids were added and the domains joined. After mutation of these domains to the human sequence and partial refinement, this model was used in a final Phaser search which weakly found the remaining two substrate-binding domains (Phaser round 3, Figure 3.12 and Table 3.7c). To find the final substrate-binding domain, clashes had to be allowed to produce a solution. While the Z-scores are low, the LLG does increase, indicating that the model is explaining the data better. The domain termini were again placed in proximity and could be joined through continuous Fo-Fc density and the clashes could be manually resolved. The Euler angles and fractional positions of the final solution are given in Table 3.7d. It should be noted that while both the *T. thermophilus* (PDB code 2CUK) and *P. horikoshii* (Yoshikawa *et al.*, 2007) GRHPRs show higher sequence identity (42% and 39% respectively) then the *H. methylovorum* GRHPR and would have made more appropriate search models, they were not available at the time.

No. monomers/AU	Matthews Coeff.	%solvent
2	5.3	76.8
3	3.5	65.2
4	2.7	53.6 *
5	2.1	42.0
6	1.8	30.4

Table 3.6 Matthews analysis for the ternary GRHPR data.

The star indicates the most likely description of the AU.

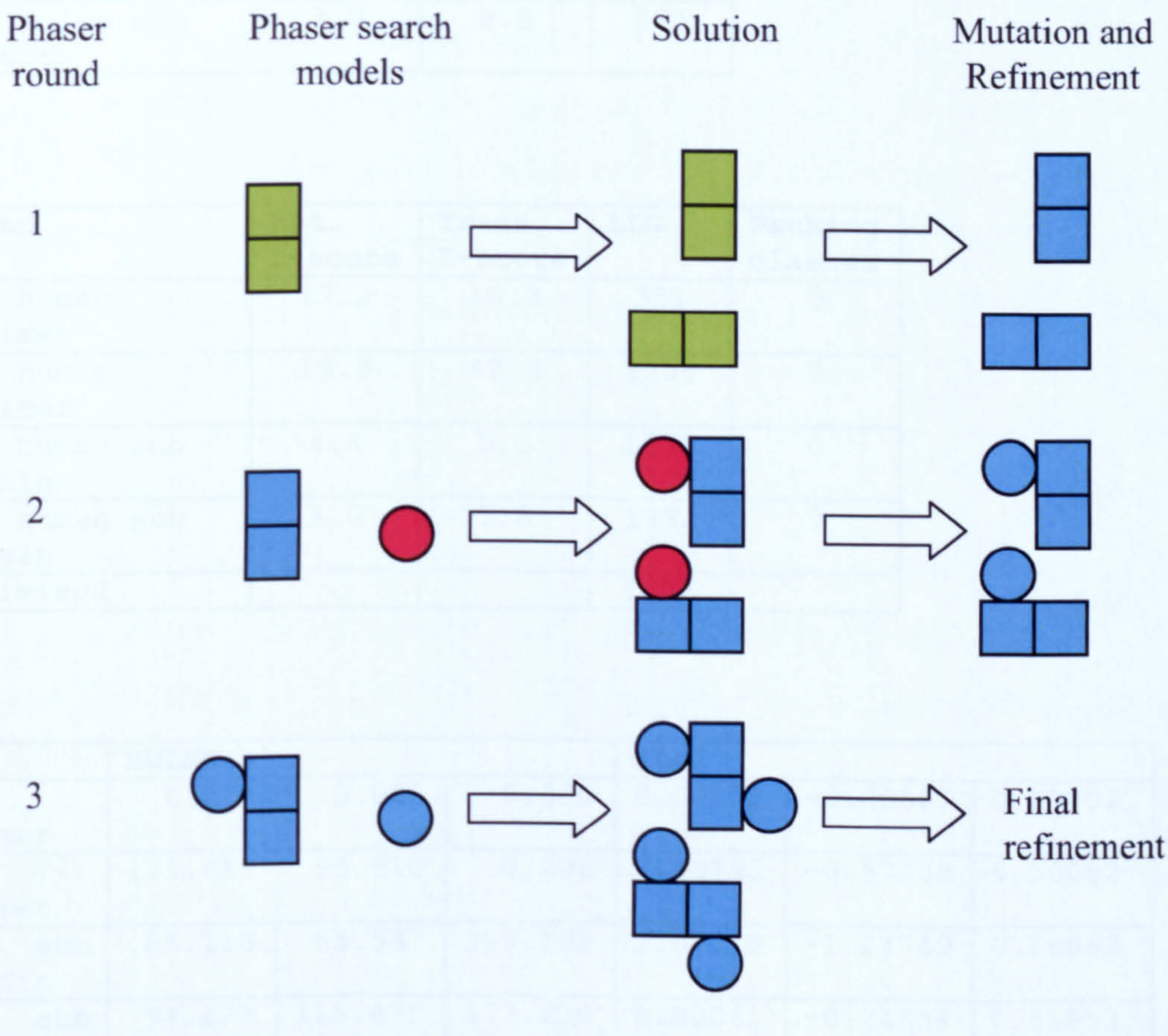


Figure 3.12 Schematic of how Phaser was used stepwise to find the full ternary GRHPR solution.

The solution at the end of each round was used to create the search model for the subsequent round. The co-enzyme binding domains (squares) were retained as a dimer, while the substrate binding domain (circle) was used as a separate search model. Domains of the *H. methylovorum* sequence are in green, those mutated to the human sequence are in blue, and the poly-alanine model is indicated in red.

a)

Found	Rot. Z-score	Trans. Z-score	LLG
1st bacterial co domain dimer	6.8	5.8	64
2nd bacterial co domain dimer	4.8	10.7	200

b)

Found	Rot. Z-score	Trans. Z-score	LLG
1st human co domain dimer	11.9	9.5	221
2nd human co domain dimer	10.8	30.3	765
1st polyA sub domain	3.5	7.0	829
2nd polyA sub domain	3.8	8.3	895

c)

Found	Rot. Z-score	Trans. Z-score	LLG	Packing clashes
1st human 'trimer'	17.2	10.3	351	0
2nd human 'trimer'	19.9	42.4	1308	0
1st human sub domain	4.4	5.2	1333	0
1st human sub domain	4.0	5.6	1382	9
Optimised			1626	

d)

	EULER			FRAC		
1st trimer	0.018	0.000	0.000	0.00001	-0.00087	0.00002
2nd trimer	179.617	95.610	0.206	0.99793	-0.57638	0.50082
1st sub domain	85.118	63.537	359.192	1.07518	-1.24749	0.16662
2nd sub domain	94.127	116.435	177.630	0.90061	-0.74854	0.31953

Table 3.7 Results of Phaser search rounds for ternary solution.

Each table gives the Z-scores for the solution after each Phaser round as displayed in Figure 3.12. Table a) round 1, b) round 2, c) round 3. Table (d) gives the Euler angles and fractional positions of the final solution.

3.2.6.2 Apo

As for the ternary data, the ScalePack output file was first converted into CCP4 mtz format and Matthews analysis performed. Again the output (Table 3.8) suggests the unit cell contains 4 monomers (2 dimers), with a Matthews coefficient of 2.7 and a solvent content of 54.5%.

The previously solved human ternary GRHPR structure was used as the molecular replacement model. Firstly, one of the dimers from the ternary asymmetric unit (chains A and B) was striped of its water, sulphate and ligand atoms. To reduce model bias in the potential domain movements, this search model was split into its separate domains. The coenzyme binding domains (108 – 297; numbering as from the human GRHPR) from both chains were retained in the dimeric form, and used as the first search ensemble. As this already consisted of the core of a dimer, 2 were expected to be found within the asymmetric unit. One substrate binding domain (from chain A, 5 - 105 and 300 - 328) was used as the second search ensemble, 4 of these were to be found within the asymmetric unit. Initial searchers in Phaser found one complete dimer and a partial dimer which was missing a single substrate binding domain (Table 3.9a). The two substrate binding domains identified with significant Z-scores (translational Z-scores of 14.8 and 19.5) together with one of the coenzyme binding domain dimers formed a complete GRHPR dimer. This was extracted from the solution file and used to search for the other dimer. Both dimers were successfully found with no packing clashes, excellent Z-scores and a very high final LLG indicating the strength of this solution (Table 3.9b). The Euler angles and fractional positions of the final solution are given in Table 3.9c.

No. monomers/AU	Matthews Coeff.	%solvent
2	5.4	77.3
3	3.6	65.9
4	2.7	54.5 *
5	2.2	43.2
6	1.8	31.8

Table 3.8 Matthews analysis for the apo GRHPR data.

The star indicates the most likely description of the AU.

a)

Found	Rot. Z-score	Trans. Z-score	LLG
1st co domain dimer	15.9	28.5	382
2nd co domain dimer	13.1	52.1	1650
1st sub domain	3.2	14.8	1763
2nd sub domain	3.4	19.5	1884
3rd sub domain	3.5	4.8	1909
Optimised			1909

b)

Found	Rot. Z-score	Trans. Z-score	LLG
1st dimer	23.6	27.4	738
2nd dimer	20.3	62.8	3393
Optimised			3393

c)

	EULER			FRAC		
1st dimer	270.06	137.858	0.201	-0.31730	-0.00223	-0.12901
2nd dimer	90.091	42.128	180.25	0.18248	0.00133	-0.12826

Table 3.9 Results of Phaser search rounds for apo solution.

Each table gives the Z-scores for the solution after each Phaser round as described in section 3.2.6.2. Table a) round 1, b) round 2. Table (c) gives the Euler angles and fractional positions of the final solution.

3.2.7 Model building and refinement

Model building was carried out using the graphics program Coot (Emsley and Cowtan, 2004), while Refmac5 (Murshudov, Vagin and Dodson, 1997) was used for refinement. Programs from CNS were also used in an attempt to help refinement but found to offer little improvement. The success of refinement was monitored by a decline in both R values. Graphs representing the successive decline of the R values after specific major refinement steps are give in Figure 3.13, for the ternary data refinement and Figure 3.14, for apo data refinement. The refinement steps are listed in Table 3.10 and Table 3.12 respectively and will be described in more detail below. There are no initial R values (Refinement stage 0) as Phaser does not calculate them. The first R values are therefore the result of some alteration to the original Phaser solution. In between each major stage in refinement the resulting model and density maps were analysed and alterations to side-chain and mainchain positions were made where indicated by the Fo-Fc difference map. The successive rounds of Phaser searches and model building used it find the final ternary solution are represented schematically in Figure 3.12.

3.2.7.1 Ternary

Inspection of the maps produced from the initial Phaser solution of 4 bacterial coenzyme domains showed little identifiable side chain density, but a majority of continuous density for the mainchain. Some manual realignment of mainchain atoms and rounds of refinement in Refmac5 were carried out using tight non-crystallographic symmetry (NCS) restraints (stage 1). With NCS restraints it is possible to manually define regions of multiple protein chains that should remain the same and not be refined to a local minimum. This is important during early refinement where the model is of low quality and there is no data to suggest the chains are different. Additionally, a high weighting was placed on the geometric restraints to prevent refinement resulting in unreasonable geometry deviations. Subsequent examination of the density maps produced by Refmac5 showed a marked improvement in quality, with some side chain density visible. Based on the sequence alignment (Figure 1.2), the model was mutated to the human GRHPR sequence. This was assisted by the presence of a number of noteworthy areas of positive Fo - Fc density, which corresponded to large hydrophobic amino acids in the human sequence not present in the bacterial. These included Trp 145, Tyr 149, Phe 230 and 240, and Tyr 255 (numbered as in human GRHPR). These

were used as set points around which the sequence could be mutated, ensuring the register was retained. Sections of one domain were mutated, the model run through Refmac5 and the resulting maps analysed for negative difference density over 'new' sequence indicating a register mistake and the appearance of new positive density. As the domains became more human-like, the density maps overall improved and more side chains could be positively identified. Finally, all 4 coenzyme binding domain had been mutated to the human sequence and there could be reasonable confidence in the sequence register (stage 2). This produced a moderate decrease in both R values. While there was a small amount of unmodelled density visible in the region where the substrate binding domains were expected, this was not sufficient to proceed with manual building. Furthermore, automatic building with ARP/wARP (Perrakis, Morris and Lamzin, 1999) was not successful. One of the human sequence coenzyme domain dimers was used in the second Phaser search.

A similar process was carried out on the two substrate binding domains found during the second Phaser search. Initial refinement of the main-chain positions (stage 3) was followed by mutation (stage 4). Here, positive density corresponding to Trp 9, Tyr 23, Phe 79 and Tyr 318 of the human sequence was used to confirm the register. The mutation of these domains resulted in a 5% decrease in both R values. Additionally, the 2 domains could be joined to their corresponding coenzyme domain. From these stages forward, NCS regions were defined separately for the different domains and restraints were applied. A 'trimer' made up of a dimer of coenzyme domains and a single substrate binding domain was extracted from this model and used, along with a separate substrate binding domain in the final Phaser search. The addition of these final domains completed the protein content of the asymmetric unit, and their subsequent refinement resulted in further large R value decreases (stage 5).

Over the next stages of refinement (6, 7 and 8) the NCS restraints were gradually lowered. This was done slowly, first allowing the side-chains to refine independently and then loosening the restraint on the main-chain atoms. As the model improved the weighting towards geometric parameters over X-ray data was decreased without deterioration in geometry. Interspaced with these changes, the resulting maps were inspected and manual side chain adjustments made. These included fixing bad rotamers and Ramachandran outliers as reported by the validation tools inbuilt in Coot.

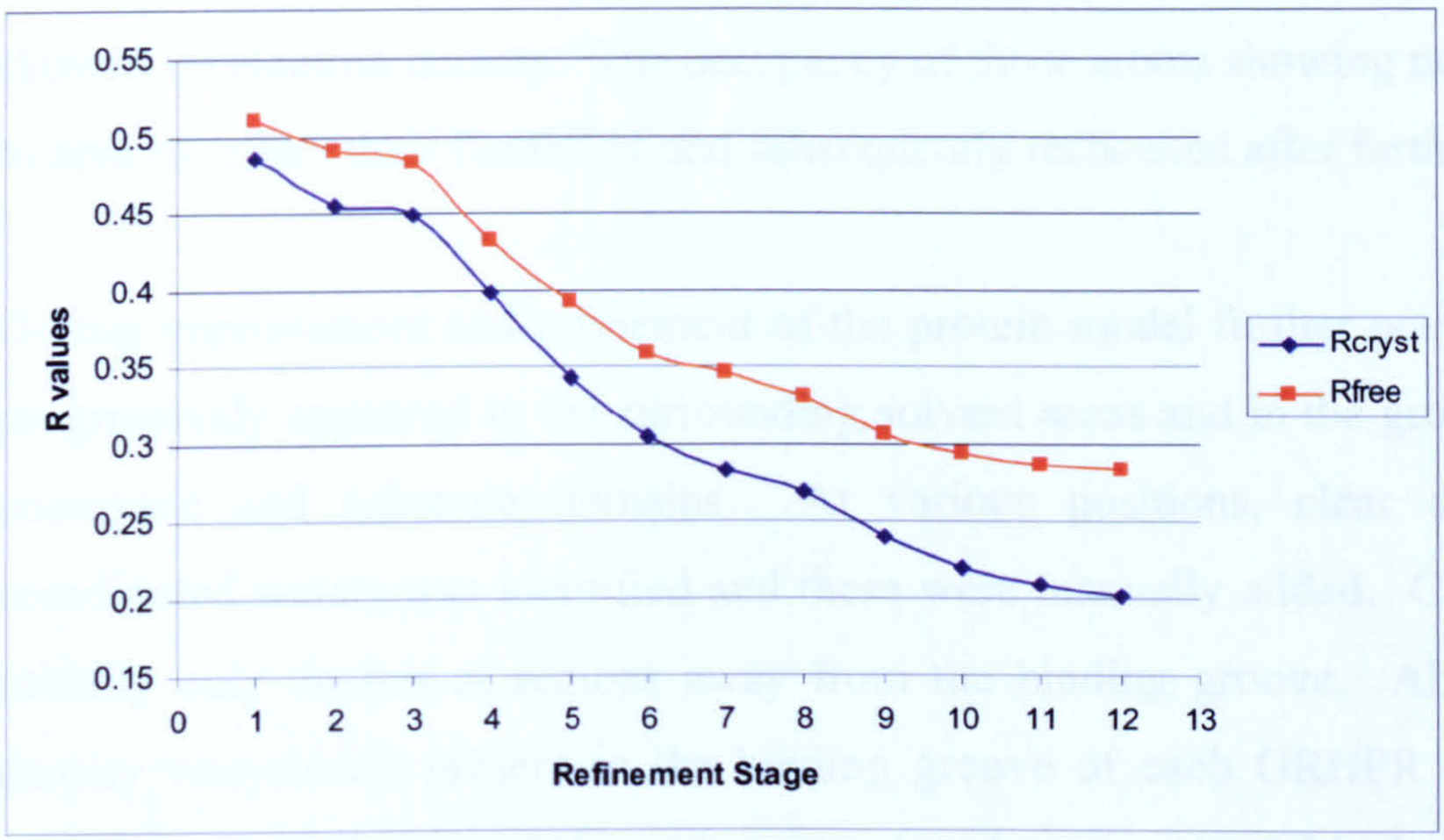


Figure 3.13 Graph of the decrease in R values over refinement of the ternary data.

R_{cryst} and R_{free} values after the key refinement steps are included and correspond to the refinement steps described in Table 3.10.

Stage	Refinement Summary
1	Tight NCS restraints and high weighting on geometric restraints
2	Mutated coenzyme domain to human sequence
3	Found 2 substrate binding domains – poly-Ala sequence
4	Mutated both substrate binding domains to human sequence
5	Found other 2 substrate binding domains – complete 2 human dimers
6	Tight mainchain NCS restraints
7	Loosening of NCS restraints
8	Removal of NCS, lowering of weighting on geometric restraints
9	Ligand addition (3 NADPH) and waters (away from binding groove; 50)
10	Ligand addition and alteration (4 NADPH at ½ occupancy, 2 oxalates at full occupancy), Coot automatic find waters (508), sulphate ions (7)
11	TLS refinement (10 TLS groups per monomer)
12	Ligand occupancy alteration, oxalate was replaced by D-glycerate and final checks of waters and side chain occupancies (610 waters, 5 sulphate ions)

Table 3.10 Major stages of refinement for the ternary data.

Summarised notes describing the key stages involved in the refinement procedure employed for the ternary GRHPR structure. Each stage corresponds to a resulting R_{cryst} and R_{free} value displayed in Figure 3.13.

Additionally, a number of flexible (ie Arg and Lys), solvent exposed side chains showed no electron density. The occupancy of those atoms showing no density was set to zero to reflect their flexibility and subsequently rechecked after further refinement.

During improvement and refinement of the protein model further positive density had progressively appeared in the surrounding solvent areas and in the groove between the coenzyme and substrate domains. At various positions, clear density for well coordinated waters was identified and these were manually added. Care was taken to initially only do this at regions away from the binding groove. Almost continuous density was clearly present in the binding groove of each GRHPR monomer in the position occupied by NADH in other complexed structures of this family. A NADPH monomer was imported into Coot and manually fitted to this density in 3 of the 4 monomers. This resulted in a significant reduction of both R_{cryst} and R_{free} (stage 9). A fourth NADPH molecule was fitted into the appropriate density of chain D and density in the substrate binding region was identified in chains A and C. This was originally modeled as oxalate as this is what was incubated with the protein prior to crystal growth. Gaps in the 2Fo - Fc map and negative density in the Fo - Fc map indicated that NADPH was not present at full occupancy and all were initially lowered to half occupancy. Further waters were identified with the Coot "find waters" option. This works by searching the asymmetric unit for unexplained density above a set sigma cutoff and automatically fitting water molecules to it. These locations were manually checked, and those waters in positions that lacked suitable coordination were deleted. Also, a number of larger positive density peaks were identified and modeled as sulphate ions from the crystallisation condition. The large decrease in R value seen at this stage (10) is likely due to the large increase in modeled waters.

This model was fed into the TLSMD server (Painter and Merritt, 2006) where the protein chains were partitioned into multiple TLS groups. Each chain was divided into 5, 10 or 15 TLS groups and the resulting files inputted into Refmac5 to undergo TLS and restrained refinement. The use of 10 TLS groups provided the best reduction in R values (stage 11). In the final stages of refinement of the model, positive difference density was noted near the O1 position of the modeled oxalate. Modeling this density with a water molecule resulted in an insufficient hydrogen bonding distance. Analysis by electrospray ionization mass spectrometry (ESI-MS) was carried out on the di-

sodium oxalate (Sigma) used in the crystal tray set up. This showed a distinct peak at a mass corresponding to hydroxypyruvate (data not shown), one substrate for GRHPR. It was thought that during the incubation period prior to crystal tray set up, the GRHPR would convert this hydroxypyruvate into D-glycerate. Consequently, the substrate density was modeled and refined with D-glycerate rather then oxalate resulting in the density being well explained (Figure 3.22b). A final analysis of the density maps was carried out, checking water positions, side chain and ligand occupancies (stage 12).

The N-terminal 6 residues, 4 from GRHPR and the 2 remaining tag amino acids, were disordered and no density was seen. These have been left out of the final model which contains two GRHPR dimers from residue 5 (chains A and C) or 6 (chains B and D) to the C-terminus at residue 328, 610 waters and 5 sulphate ions. The presence of different ligands and their occupancy depended on which protein chain they were associated with and are described in Table 3.11. In summary, the final model contained two GRHPR dimers related to each other by an approximately 90° rotation along their short axis. Interdimeric contacts occur via ordered water molecules between coenzyme binding domains. Each dimer consists of a ternary (NADPH and D-glycerate bound) monomer and a binary (NADPH bound) monomer (Figure 3.17) and resulted in a final R_{cryst} of 20.1% and R_{free} of 28.2%. The final refinement statistics are given in Table 3.13. It should be noted that the two binary chains (B and D) correlate to the substrate binding domains which could only be found during the final Phaser search. A detailed description of the structure is given in the discussion and the quality of the electron density maps can be seen in Figure 3.22b.

Chain	NADPH	Occupancy	D-glycerate	Occupancy	State
A	Yes	0.75	Yes	0.75	Ternary
B	Yes	0.50	No	n/a	Binary
C	Yes	0.75	Yes	0.75	Ternary
D	Yes	0.50	No	n/a	Binary

Table 3.11 Presence of ligands for each chain in the "ternary" asymmetric unit of GRHPR.

3.2.7.2 Apo

Unlike the ternary solution, here the molecular replacement model and the target structure were the same protein, resulting in refinement starting at a more advanced stage. The density maps produced directly from Phaser were largely complete, including side chain density. Initial analysis of these maps and the domain positions were carried out in Coot (Emsley and Cowtan, 2004). Continuous difference density was present making it possible to add in residues 106, 107, 298 and 299, connecting the coenzyme binding domains to their substrate binding domains. Additionally, at this stage it was noted that the 217-223 loop region in all 4 protein chains had very little electron density. This region was deleted from subsequent refinement to prevent it contributing model bias (stage 1). Tight NCS and high geometry restraints were also used. Subsequent model alterations were carried out, including changing the occupancy of flexible side chains and some minimal main chain movements. Loosening of the NCS restraints resulted in a significant drop in both R_{cryst} and R_{free} of 2-3% (stage 2). However their complete removal (stage 3) only produced a minimal effect. Analysis of the model with the MolProbity web server (Davis *et al.*, 2007) at this point indicated a large number of bad rotamers were present. Manual improvement of these rotamers, followed by refinement in Refmac5 did not resolve the problem. This was likely due to Refmac5 forcing bad geometry in an effort to improve other statistics and is due to the poorer quality data in many of these regions. Even though the overall geometry weighting was high, it was found that manually tightening the torsion restraints, from 1.0 to 5.0, within Refmac5 resulted in an almost immediate improvement to the rotamers. Importantly, this brought R_{free} below 30%, while the corresponding increase in R_{cryst} suggests some amount of model overfitting had been eliminated (stage 4). Additionally, a solvent exposed region of the substrate binding domain of chain D (residues 45 - 53, 61 - 74, 88 - 92, and 96 - 100), distant from crystal contact sites, showed very poor density and the occupancy of the region was reduced. As this region is well ordered in all other chains, tight NCS restraints between it and the same region in chain B were used to force it to retain a 'likely' structure. The use of 10 TLS groups per monomer during refinement resulted in a further 1% decrease in both R values (stage 5). Interestingly, this had improved the model to a point where the extra torsion restraint could be reduced back to the default setting without a negative effect. Only 48 waters could be identified and fitted to density compared to the 610 of the ternary structure. This lower number is due to the lower resolution of

this data set (2.85Å). Two additional areas of positive solvent density too large to be water molecules were modeled as Mg^{2+} ions, which were present as the crystallisation salt (stage 6). No significant difference density had appeared for the 217 - 223 loop region during the course of refinement and it remained deleted. Finally refinement resulted in the reasonable R_{cryst} value of 21.7% and an R_{free} of 27.9%. Further refinement statistics are given in Table 3.13.

In summary, the final model contained two GRHPR dimers, one 'horizontal' and the other 'vertical'. Within the asymmetric unit, interdimeric contacts occur via direct hydrogen bonds largely between a substrate binding domain in one dimer and a coenzyme binding domain in the other. Each monomer consists of residues 6 – 327 (or 325 in chain D) and lacked density for the coenzyme binding loop 217 - 223, which is not present in the final model. This loop is discussed further in section 3.3.2.3. No significant density was present in the binding grove in any of the 4 monomers.

3.2.8 Validation and deposition

Once refinement had converged and the R-factors stabilised, the structure was validated by analysis in SFcheck (Vaguine, Richelle and Wodak, 1999) and/or MolProbity (using the all-hydrogen option, Davis *et al.*, 2007). Identified Ramachandran outliers were examined and where possible fixed, as were excessive clashes and bad rotamers. The structures were re-refined after any changes. The Ramachandran plots from SFcheck (ternary) or MolProbity (apo) are shown below (Figure 3.15 and Figure 3.16). The final ternary structure coordinates and structure factors were deposited into the Protein Data Bank (PDB) on 14/03/2006 and released on 18/07/2006 under the code 2GCG. This structure has also been published (Booth *et al.*, 2006). The final models are discussed in the following section and are represented in Figure 3.17. It should be noted that after this deposition, a structure of human GRHPR was deposited from the Center for Eukaryotic Structural Genomics (unpublished, PDB code 2H1S). Despite this structure being modeled in the apo state it contains extensive density within the active site cleft and has highly similar (rmsd 0.33Å over 321 equivalent Cα positions) domain positions to the ternary data produced in this thesis. Because of the uncertain state of this enzyme it shall not be discussed further.

The hinge region between the two domains lies at residues 104 – 107 and 298 – 301 which coincides with the only significant Ramachandran outlier of the ternary model; Asp105 (Figure 3.15). This well defined residue is within the disallowed region in all 4 chains and suggests a deformation of this residue has occurred to facilitate the present domain positions. All other significant outliers in both models occur in areas (loop regions) of poor density.

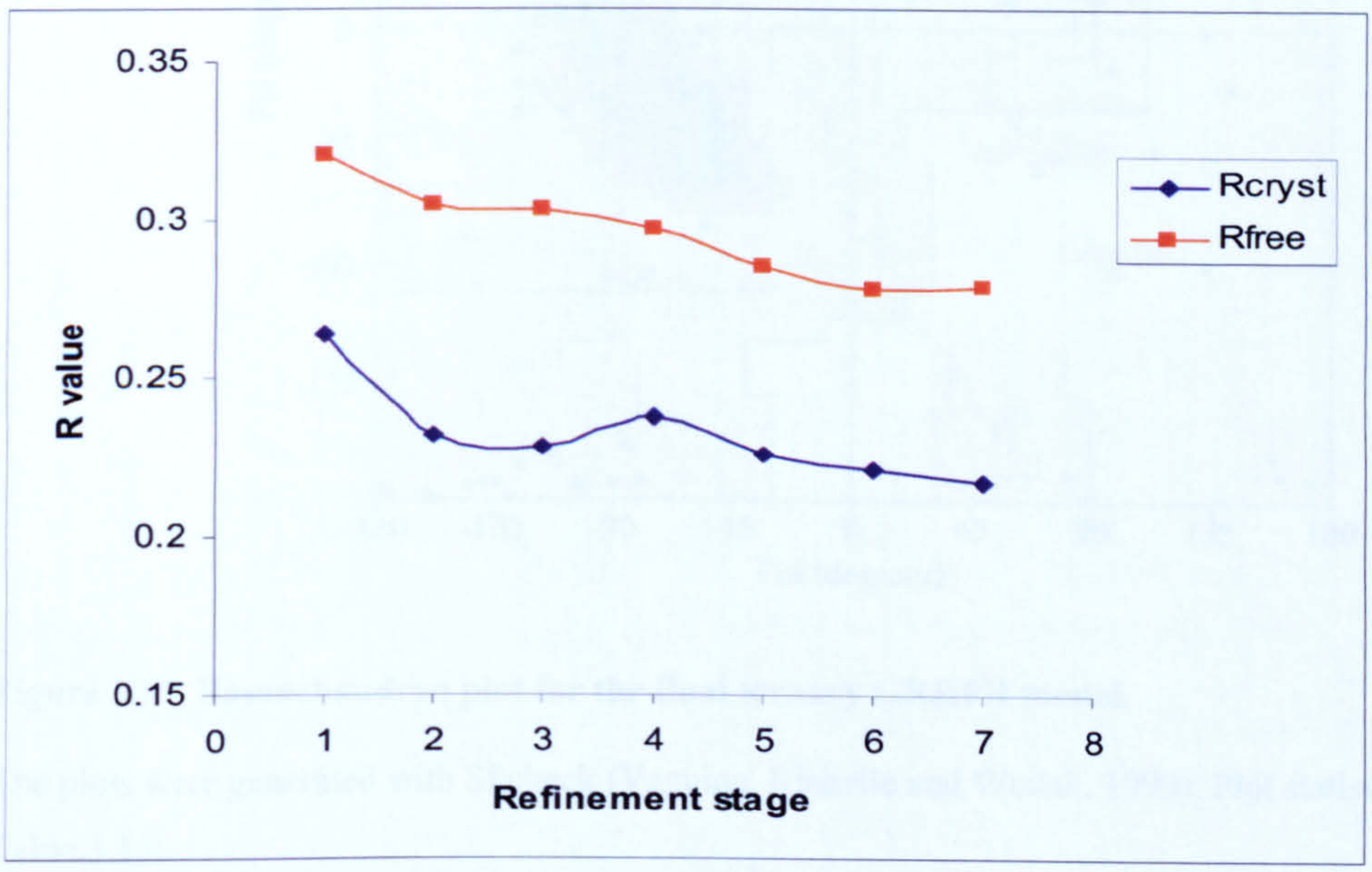


Figure 3.14 Graph of the decrease in R values over refinement of the apo GRHPR data.

R_{cryst} and R_{free} values after the key refinement steps are included and correspond to the refinement steps described in Table 3.12.

Stage	Refinement Summary
1	Tight mainchain NCS restraints and high weighting on geometric restraints
2	Loosening of NCS restraints
3	Removal of NCS
4	Tightened torsion restraints to fix bad rotamers, NCS restraints on specific regions
5	TLS refinement (10 TLS groups per monomer)
6	Reduction of torsion restraint to default, 48 H_2O and $2Mg^{2+}$
7	Final checks of waters and side chain occupancies

Table 3.12 Major stages of refinement for the apo GRHPR data.

Summarised notes describing the key stages involved in the refinement procedure employed for the apo GRHPR structure. Each stage corresponds to a resulting R_{cryst} and R_{free} value displayed in Figure 3.14.

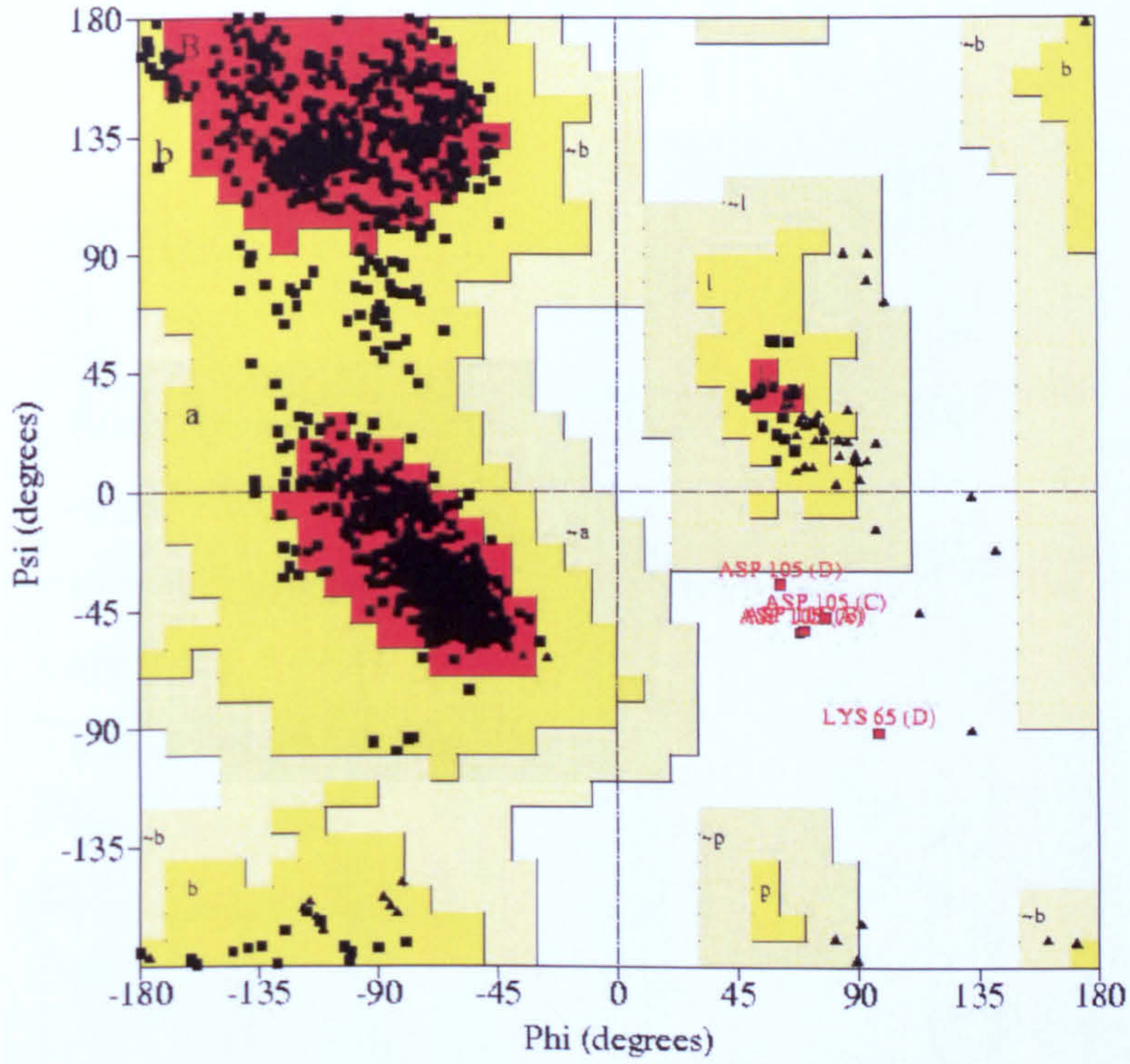


Figure 3.15 Ramachandran plot for the final ternary GRHPR model.

The plots were generated with SFcheck (Vaguine, Richelle and Wodak, 1999). Plot statistics are given in Table 3.13.

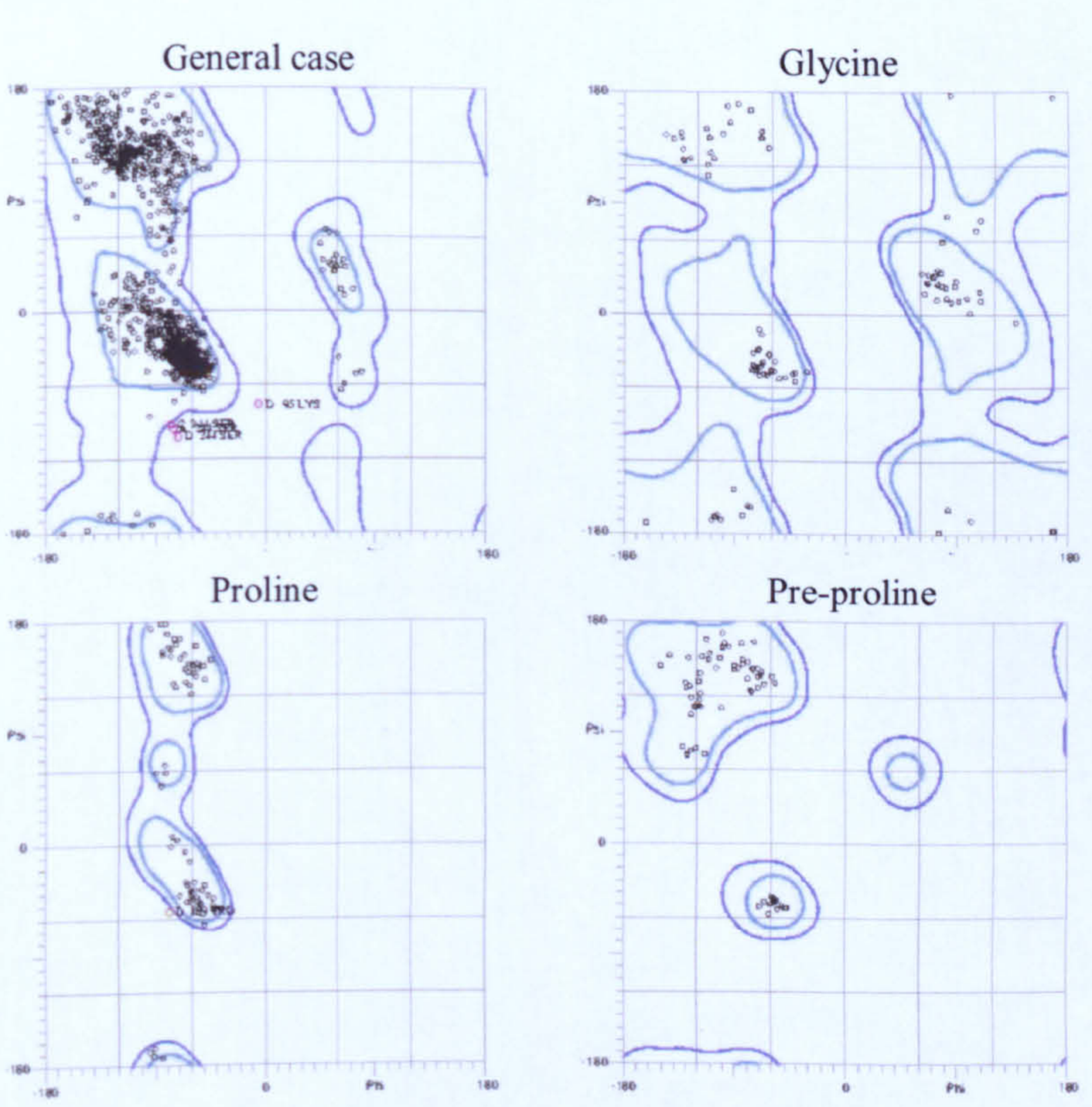


Figure 3.16 Ramachandran plots for the final apo GRHPR model.

The plots were generated with MolProbity (Davis *et al.*, 2007). Plot statistics are given in Table 3.13.

	Ternary	**Apo**
R_{cryst} (%)	20.1	21.7
R_{free} (%)	28.2	27.9
RMS deviation for bond length (Å)	0.017	0.007
RMS deviation of bond angle (°)	1.866	1.073
Ramachandran plot (%)		
Most favoured	88.2	96.0
Additionally allowed	11.3	3.6
Disallowed	0.4	0.4

Table 3.13 Final refinement statistics from both ternary and apo GRHPR data.

The R_{cryst} and R_{free} values indicate a reasonable final model. The good geometry of the model is indicated by the low RMS deviation values and low percentage of residues in the disallowed region of the Ramachandran plots.

3.3 Discussion

The work presented in the preceding sections details the solution of the first structures of human glyoxylate reductase/hydroxypyruvate reductase and the first true ternary structure of a D-2-hydroxyacid dehydrogenase family member. Structures were solved by X-ray crystallography in an attempt to provide information on the three major states (ternary, binary and apo) of this enzyme. Fortuitously, one crystal structure provided information for both the ternary and binary states, with a second being solely apo.

In the following, the general GRHPR structure will be described and differences between the three states discussed. The details of both coenzyme and substrate binding will be described, the latter having added importance as it is the first time the structure of a true ternary complex has been observed for this enzyme family. Where appropriate comparisons to other family members will be made and finally, a structural explanation will be provided for why observed mutations give rise to PH2.

3.3.1 General structure

The dimeric nature of human GRHPR, as expected in this enzyme family and indicated earlier by the gel filtration results, is apparent in the crystal structures where two dimers are present in each asymmetric unit. Each dimer is formed by a staggered, "back-to-back" arrangement of monomers, forming an elongated ellipsoid with dimensions of approximately 100Å x 57Å x 44Å and displaying 2-fold rotational symmetry (Figure 3.17a). Each monomer has a dumb-bell shape, consisting of two compact domains separated by a deep cleft which contains the coenzyme binding site and, at the bottom, the catalytic site and hinge region (Figure 3.17b). This is the expected D-2-hydroxyacid dehydrogenase fold as seen in previous structures from this family (Figure 1.3).

Both domains are α/β/α structures and represent variations on the Rossmann nucleotide-binding fold (Rao and Rossmann, 1973). The smaller substrate binding domain comprises residues 5 (the first 4 residues are disordered in all crystal structures) to 106 and 299 to 328 (the C-terminus). This domain includes a central 5 stranded parallel β-sheet which runs perpendicular to the binding cleft flanked on one side by 2 helices and 3 on the other (Figure 3.17b and Figure 3.18).

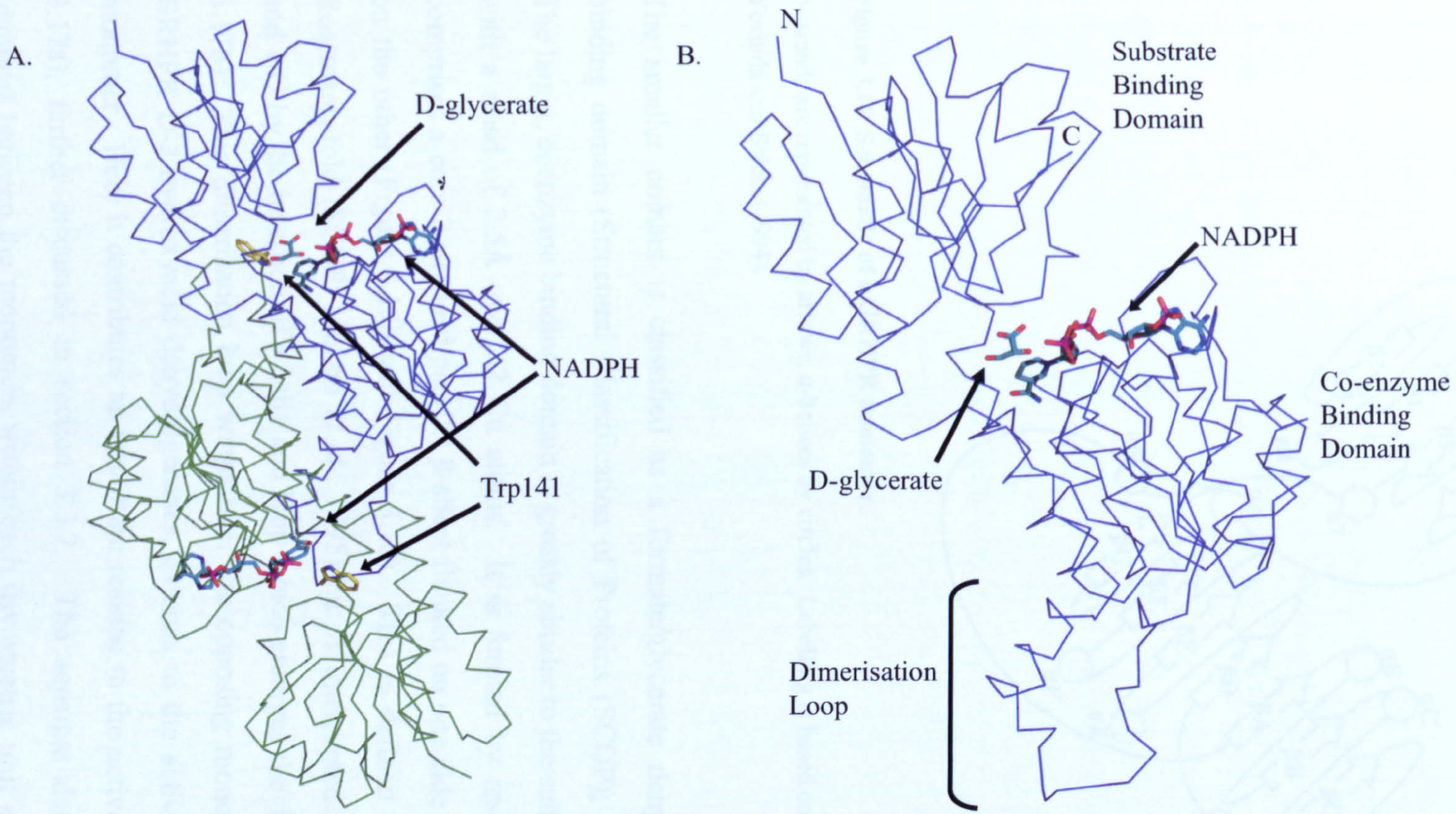


Figure 3.17 Overall structure of human GRHPR.

A) Cα trace of a one of the dimers present in the ternary asymmetric unit. Subunit A shown in blue, subunit B in green. The Trp141 sidechain is shown as stick representation in yellow and labelled; and the co-enzyme and substrate are in cyan and also labelled. The star indicates the 217 – 223 loop which is disordered in the apo GRHPR subunits. B) Cα trace of a single ternary subunit, subunit A. Co-enzyme and substrate are in cyan and labelled. The amino and carboxyl termini are labelled N and C, respectively. The individual substrate and co-enzyme binding domains are indicated, as is the dimerisation loop.

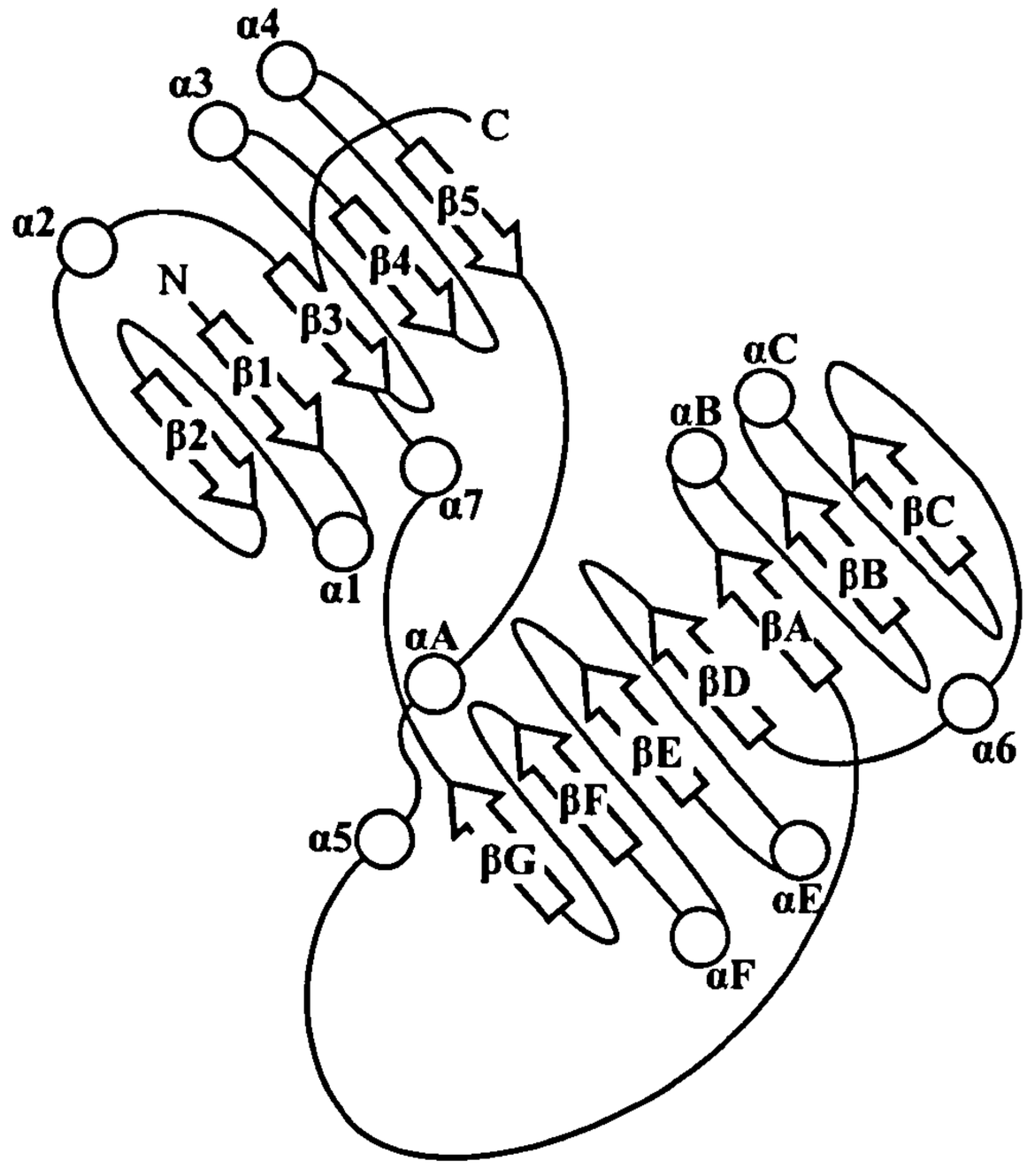


Figure 3.18 Schematic of a GRHPR monomer.

β-strands are represented by arrows, α-helices by circles. Labelling is based on that of Goldberg, Yoshida and Brick, (1994).

The smaller domain is classified as a formate/glycerate dehydrogenase substrate-binding domain (Structural Classification of Proteins (SCOP); Murzin *et al.*, 1995). The larger, coenzyme binding domain is grossly similar to the substrate binding domain with a rmsd of 2.5Å over 62 Cα atoms. It is formed by residues 107 – 298 and comprises a core 7 stranded parallel β-sheet flanked on one side by 3 helices and by 4 on the other (Figure 3.17b and Figure 3.18). This is defined as a NAD(P)-binding Rossmann-fold (SCOP; Murzin *et al.*, 1995) but is interrupted between residues 123 and 149 by the insertion of a helix and loop which enlarges the dimer interface (Figure 3.17b). This dimerisation loop wraps over the opposing monomer and, unlike non-GRHPR D-2-hydroxyacid dehydrogenases, extends to the active site of the opposite monomer. Here it contributes an important residue to the active site, Trp141 (Figure 3.17a), further discussed in section 3.3.7. The separate domains are effectively identical between the monomers within each asymmetric unit and in the three states represented in the crystal structures, with rmsd values of 0.3Å – 0.4Å (131 Cα) between the substrate binding domains and 0.3Å – 0.6Å (179 Cα) between coenzyme binding domains.

Dimerisation has resulted in the burying of 3161Å (19%) of solvent-accessible surface area between monomers, predominantly due to interactions between the coenzyme binding domains. This is very similar to the area buried in the *H. methylovorum* GRHPR structure (3100Å; Goldberg, Yoshida and Brick, 1994), but significantly more then that buried in the related D-2-hydroyisocaproate dehydrogenase dimer (2365Å, 14.5%; Dengler *et al.*, 1997), a difference likely to be due to the extended dimerisation loop present in the GRHPR family.

As mentioned above, the overall fold is as expected for the D-2-hydroxyacid dehydrogenase family, with the coenzyme binding domain particularly homologous to that of other family members (rmsds 0.9Å – 1.6Å: Table 3.14). This is unsurprising as all of these domains bind either NADH or NADPH ligands and therefore need to maintain a compatible binding site. The greater variation seen between the substrate binding domains (rmsd values of 1.1Å – 2.5Å: Table 3.14) is likely to arise from the weaker constraints in this domain which binds a range of substrates across the family.

Structurally, the domains of human GRHPR is seen to be most similar to *T. thermophilus* GRHPR (unpublished, PDB code: 2CUK) of the known equivalent structures (Table 3.14), despite *P. horikoshii* GRHPR (Yoshikawa *et al.*, 2007) having a slightly higher sequence identity (42% verses 39%). However, the latter has a small loop insertion in the coenzyme binding domain and a short loop deletion in the substrate binding domain which may explain this discrepancy.

3.3.2 Comparisons between states

3.3.2.1 Domain position

While the structure of individual domains changes little between the enzymatic states, the spatial relationship between domains has been seen to vary in other D-2-hydroxyacid dehydrogenases when different states have been compared (Lamzin *et al.*, 1994; Razeto *et al.*, 2002). Relative domain movements are facilitated by the hinge regions centered on residues 106 and 299 (human GRHPR numbering) and allow the enzymes to adopt 'open' and 'closed' conformations. These movements allow substrate access to the active site in the 'open' conformation and subsequently shield it

from solvent in the 'closed' conformation. Despite the "ternary" crystal structure containing both ternary and binary monomers there is little variation in the relative positions of the substrate binding and coenzyme binding domains (Figure 3.19a and Table 3.15). This overall similarity is reflected in the active site where a high degree of correlation between residues is apparent when the structures are superimposed (Figure 3.23a). The two ternary monomer chains (A and C) are effectively identical (Table 3.15) with the most divergent monomer (D; binary) representing the most 'open' conformation in comparison to the 'closed' conformation of the ternary chains. This monomer displayed the poorest quality electron density for bound coenzyme and has a substrate binding domain with average temperature (*B*) factors significantly higher then those of the other monomers ($B_{average}$ = 43.8Å^2 versus 30.0Å^2). This suggests flexibility in the position of this domain which is perhaps partly restrained in the structure due to crystal packing restraints. Additionally, investigation of an atom space filling model (in CCP4mg; Potterton *et al.*, 2002) indicates that despite being the most 'open' binary state there is insufficient room for subsequent substrate binding. This is likely to therefore represent a partially closed conformation that has been 'selected' by the crystal lattice.

Greater variation in relative domain positions is observed when the individual apo GRHPR monomers are compared (Figure 3.19b, Figure 3.20 and Table 3.16). This implies there is greater inherent flexibility in the unliganded enzyme. Luzzati plots (SfCheck; Vaguine, Richelle and Wodak, 1999) indicate an average coordinate error of 0.40 Å in the apo structure compared to 0.30 Å in the ternary structure. This provides an indication of the upper limit of error in atomic positions in the models: the rmsd values in Table 3.17 generally exceed these errors indicating that the displacements observed are likely to be real. The packing restrictions of the similar crystal lattices, obtained under similar crystallisation conditions (Figure 3.7), would be expected to restrain movement between the domains and hence reduce the observable variation. Determination of the crystal structure of the apo state under highly divergent crystallisation conditions might show a greater variation in the inter-domain arrangement, although no satisfactory data from alternative crystals was ever produced during this work. It is worth noting however that D-LDH produced two distinctly different conformations under identical crystallisation conditions (Razeto *et al.*, 2002).

		Coenzyme binding domain		Substrate binding domain		Ternary monomer			Apo monomer
Structure	**Enzyme State**	**No. of equivalent Cα positions**	**RMS (Å)**	**No. of equivalent Cα positions**	**RMS (Å)**	**No. of equivalent Cα positions**	**RMS (Å)**	**Inter-domain angle[a]**	**RMS (Å)**
D-Glycerate dehydrogenase[1]	Apo	187	1.3	125	2.0	294	2.4	18.5°	2.0
T. t Glyoxylate reductase[2]	Apo	184	1.0	125	1.1	308	1.6	11.7°	1.2
P. h Glyoxylate reductase[3]	Holo + sulphate	187	1.1	128	1.3	317	1.2		1.5
D-Lactate dehydrogenase[4]	Binary	172	1.6	124	2.2	273	2.7		2.4
D-Lactate dehydrogenase[5]	Holo + sulphate	178	1.5	123	2.3	290	2.1	15.4°	2.2
D-2-hydroxyisocaproate dehydrogenase[6]	Ternary (apo confor.)	171	1.4	121	2.5	278	2.9		2.5
Transcription corepressor CtBP[7]	Ternary	177	1.5	119	1.8	302	1.8		2.1
Formate dehydrogenase[8]	Apo	184	1.6	121	2.1	260	1.9		2.0
Formate dehydrogenase[9]	Holo + azide	184	1.6	123	2.1	261	2.1	8.9°	2.5
D-Hydroxyglutarate dehydrogenase[10]	Apo	167	1.5	115	2.2	286	2.7		2.3
Phosphoglycerate dehydrogenase[11]	Binary	178	1.5	123	1.5	283	2.8		2.4

Table 3.14 Comparison of human GRHPR ternary and apo monomers with other D-2-hydroxyacid dehydrogenases.

Table 3.14 (Legend) [1]Goldberg, Yoshida and Brick, 1994; [2]*T. thermophilus*, 2CUK; [3]Yoshikawa *et al.*, 2007; [4]Razeto *et al.*, 2002, chain A; [5]Razeto *et al.*, 2002, chain B; [6]Dengler *et al.*, 1997; [7]Kumar *et al.*, 2002; [8] and [9]Lamzin *et al.*, 1994; [10]Martins *et al.*, 2004; [11]Schuller, Grant and Banaszak, 1995. [a]Angle required to overlay substrate binding domains of 'apo' and 'holo/ternary' structures after pre-alignment of the two co-enzyme binding domains. This was calculated using DynDom (Hayward and Lee, 2002), rmsd values are in angstroms (Å) and were calculated with CCP4mg (Potterton *et al.*, 2002).

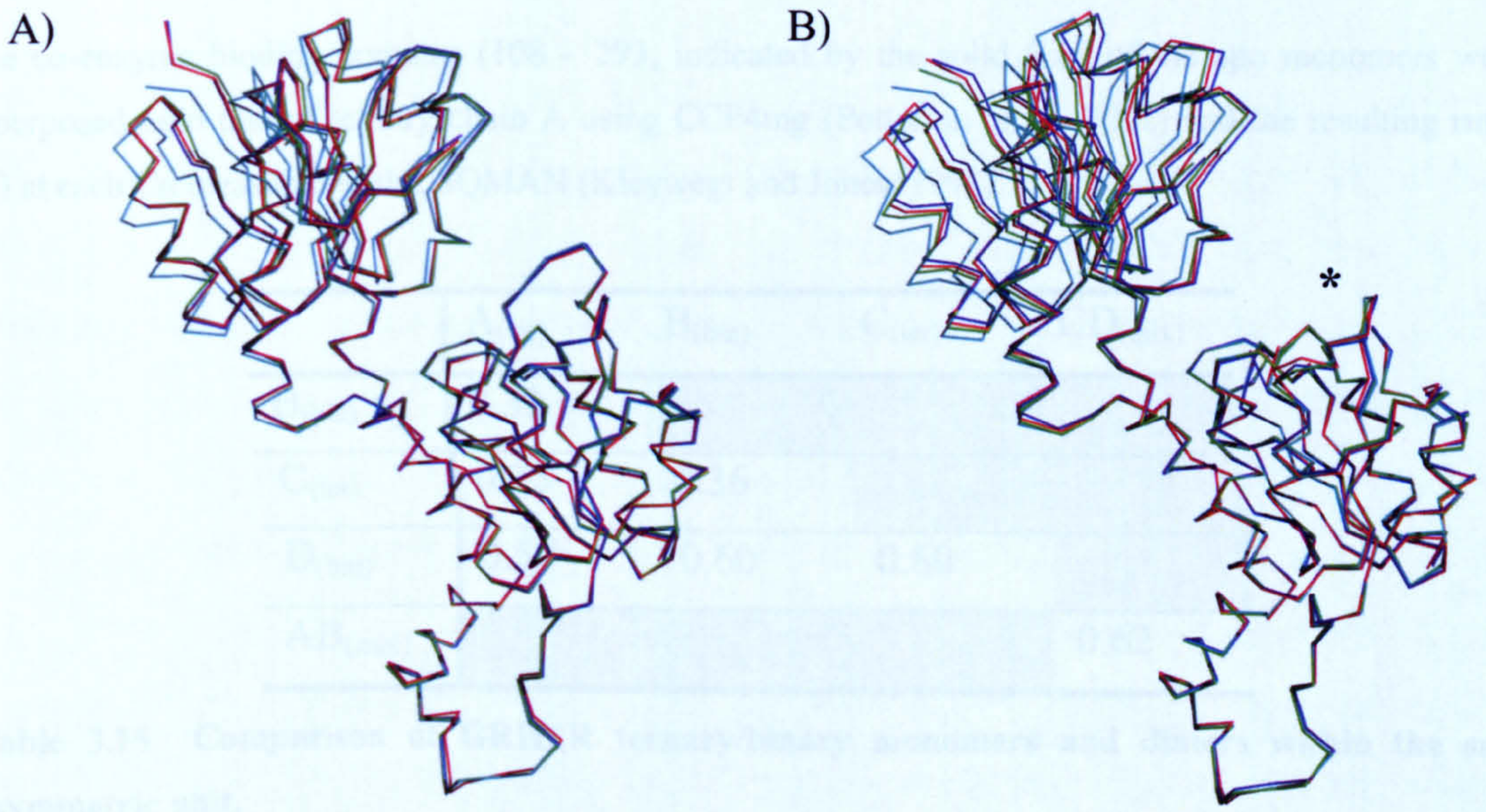


Figure 3.19 Overlay of GRHPR monomers.

A) Cα trace of the 4 subunits present in the ternary asymmetric unit. The ternary subunits are A (blue) and C (red), while the binary subunits are B (green) and D (cyan). B) Cα trace of the 4 subunits present in the apo asymmetric unit, subunits are coloured the same as in (A). The position of the 217 - 223 loop disordered in the apo structures is indicated by a star.

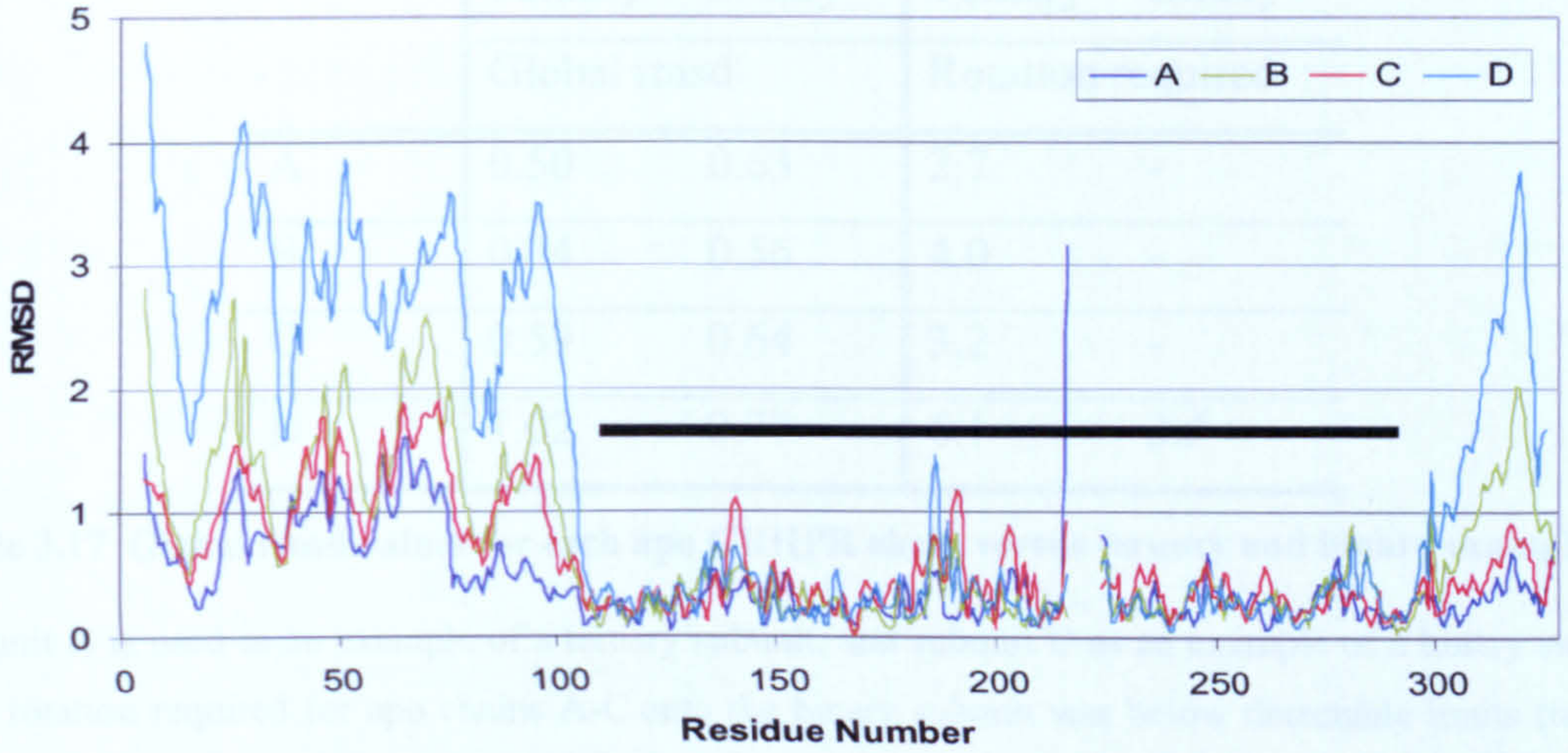


Figure 3.20 Displacement of the apo substrate binding domain.

The co-enzyme binding domains (108 – 293, indicated by the solid line) of the apo monomers were superposed onto that of ternary chain A using CCP4mg (Potterton *et al.*, 2002) and the resulting rmsd (Å) at each Cα measured with LSQMAN (Kleywegt and Jones, 1994).

	$A_{(ter)}$	$B_{(bin)}$	$C_{(ter)}$	$CD_{(mix)}$
$B_{(bin)}$	0.33			
$C_{(ter)}$	0.25	0.36		
$D_{(bin)}$	0.57	0.60	0.60	
$AB_{(mix)}$				0.62

Table 3.15 Comparison of GRHPR ternary/binary monomers and dimers within the same asymmetric unit.

The rmsd values in Å for a comparison of 323 equivalent Cα positions between each monomer.

	$A_{(apo)}$	$B_{(apo)}$	$C_{(apo)}$	$CD_{(apo)}$
$B_{(apo)}$	0.48			
$C_{(apo)}$	0.37	0.49		
$D_{(apo)}$	0.83	0.60	0.70	
$AB_{(apo)}$				0.55

Table 3.16 Comparison of GRHPR apo monomers and dimers within the same asymmetric unit.

The rmsd values in Å for a comparison of 311 equivalent Cα positions between each monomer.

	Ternary	Binary	Ternary	Binary
	Global rmsd		Rotation required	
A	0.50	0.63	2.7	-
B	0.64	0.56	4.0	-
C	0.59	0.64	3.2	-
D	1.02	0.77	6.1	3.5

Table 3.17 Global rmsd values for each apo GRHPR chain versus ternary and binary examples.

Subunit A is used as an example of a ternary subunit, and subunit D as an example of a binary subunit. The rotation required for apo chains A-C onto the binary subunit was below detectable limits (ratio of interdomain to intradomain displacement must be <1; Hayward and Berendsen, 1998). RMSD values are in Å and rotation required is in degrees.

By superimposing two coenzyme binding domains it is possible to measure the rotation required to map one substrate binding domain onto the other, giving a measure of how 'open' the binding cleft is. This was done using the DynDom server (www.sys.uea.ac.uk/dyndom; Hayward and Berendsen, 1998; Hayward and Lee, 2002) and comparisons across the 3 different states show the rotation required to go from apo to ternary vary between 2.7° and 6.1° depending on the apo monomer used (Table 3.17). For three of the apo chains the movements are small and only marginally exceed the Luzatti errors of the structures, whereas a significant change is observed for apo chain D which represents the most 'open' human GRHPR structure. This can be seen graphically in Figure 3.20 (and Figure 3.21) where equivalent positions in the substrate binding domain of apo chain D are displaced on average 2.5Å from their positions in the ternary structure, while the other chains show a displacement of around 1.0Å. This is a smaller amount than is apparent for apo *H. methylovorum* GRHPR (approximately 6.5Å, with a 18.5° rotation; Table 3.14 and Figure 3.21). A restricted domain conformation has previously been seen in the related D-2-hydroxisocaproate dehydrogenase (D-HicDH) structure which, despite binding both coenzyme and substrate, unexpectedly remains in a fully 'open' conformation (Dengler *et al.*, 1997). Here it was suggested that crystal contacts constrained the conformation that could be accommodated. Subsequently, both the open and closed domain conformations were observed in the same asymmetric unit for D-LDH (Razeto *et al.*, 2002). These observations lead Razeto *et al.*, (2002) to propose that the energetic barrier between the two conformations (open and closed) is low allowing the weak crystal contacts to

effectively compete and dominate the overall structure, a proposal supported by the relatively closed conformation of the GRHPR apo monomers presented here (Figure 3.21 and Figure 3.23b). Complete closure does however, appear to require both substrate and coenzyme ligands as neither the apo nor binary monomers are as closed as the ternary subunits (Table 3.15 and Table 3.17). The formation of multiple hydrogen bonds between the substrate binding domain and the coenzyme binding domain appears to be facilitated by binding of the substrate (described in following sections) and would be expected to stabilise this fully closed conformation. The altered spatial relationship between domains in the different states contributes to the entire ternary human GRHPR monomer most closely matching the recently determined holo (closed) *P. horikoshii* GRHPR structure (Yoshikawa *et al.*, 2007, rmsd 1.2Å) whereas the apo human GRHPR monomer is most similar to the apo (open) *T. thermophilus* GRHPR structure (rmsd 1.2Å) (Table 3.14). In general terms, the human apo state shows lower rmsd values against other apo conformations then the human ternary state, and vice versa (Table 3.14). The exception to this is the apo formate dehydrogenase, which is minimally more like the human ternary state (rmsd 1.9Å) then the human apo (rmsd 2.0Å) (Table 3.14). This is likely to be due to the relatively small domain movement seen between the formate dehydrogenase states (Lamzin *et al.*, 1994). However, it should be noted that these are generally small changes overall and care needs to be taken in comparing these crystal structures that have been determined to different resolutions and at different times.

Rigid domain movements have previously been noted when comparing apo and holo states of enzymes from this family. Formate dehydrogenase, for example, has a 8.9° rigid body rotation of the substrate binding domain towards the coenzyme binding domain occurs between the apo and holo (NADH + azide) states (Lamzin *et al.*, 1994). This is significantly smaller then the 15.4° rotation seen between the binary and holo (NADH + sulphate) structures of D-LDH (Razeto *et al.*, 2002). This larger rotation has been thought to be required for the larger substrates utilised be D-LDH in comparison to FDH (Razeto *et al.*, 2002). The rotation seen between the FDH structures is comparable to the largest rotation seen between human GRHPR apo and ternary monomers (6.1°; Table 3.17 and Figure 3.21). A low energetic barrier between open and closed conformations could be an alternative explanation for the small FDH

rotation, with the apo structure not representing a completely open state but a midway conformation selected by crystal packing restraints.

Due to the similar substrate size between GRHPR and LDH it might be expected that a comparable degree of opening of the binding cleft is required. A comparison between the human ternary GRHPR structure and the two available bacterial apo GRHPR structures, *H. methylovorum* (Goldberg, Yoshida and Brick, 1994) and *T. thermophilus* (PDB code 2CUK), indicates a required closing rotation of 18.5° and 11.7° respectively (Table 3.14), comparable to that seen for D-LDH (Razeto *et al.*, 2002). This range gives an indication of the potential variations that might be possible in the absence of the restraints of a crystal lattice. Interpretation of crystallographic structures requires a recognition of the non-physiological environment in which the protein has be crystallised and any packing constraints that may have been imposed. Consequently, individual structures should be treated with a certain level of caution. In this case, the ability to compare structures across a range of family members in various states has allowed aberrant conformations to be identified and also suggests there is a low energetic barrier between open and closed conformations.

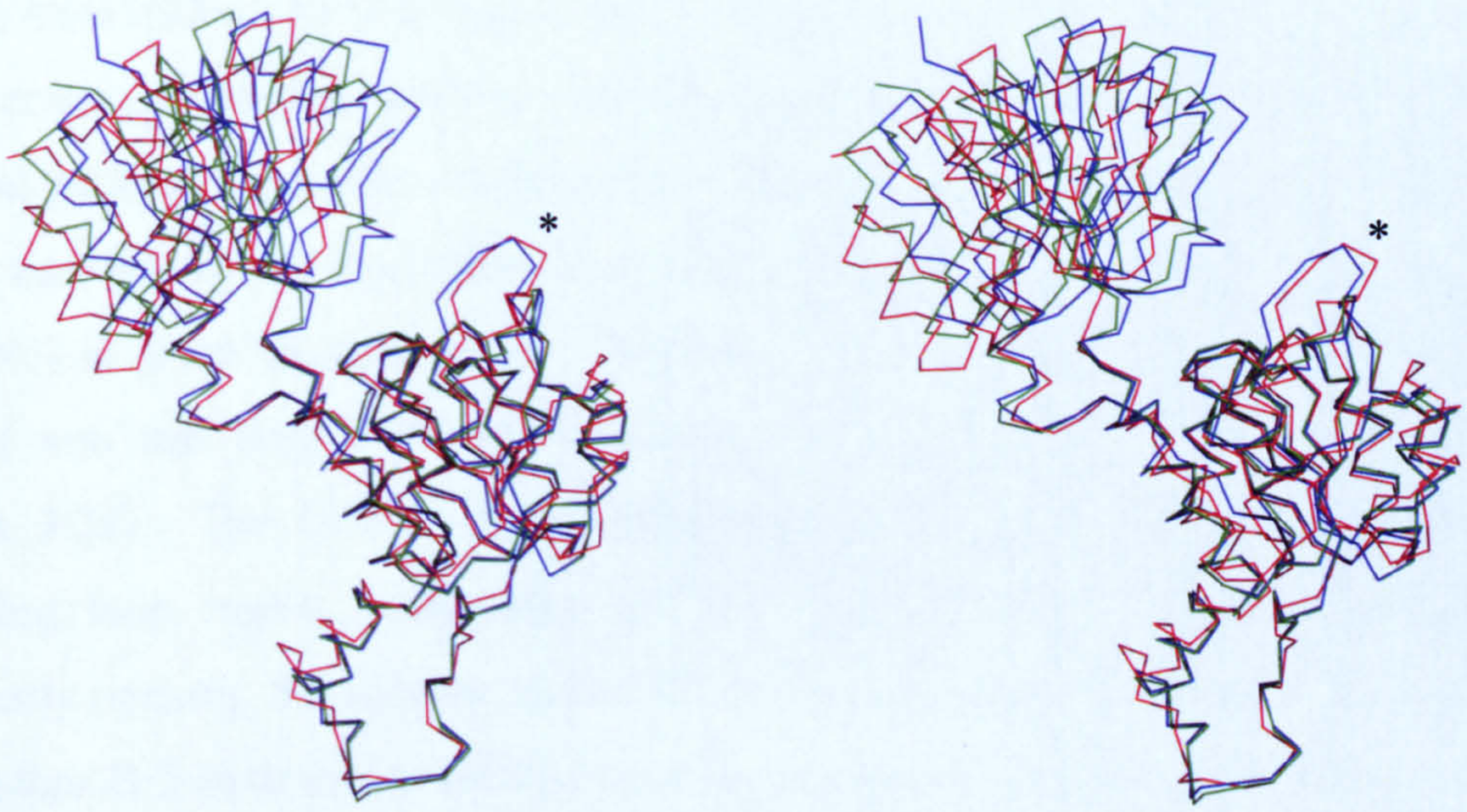


Figure 3.21 Stereo view of domain rotation.

Comparison between the human GRHPR ternary (subunit A; blue) and apo (subunit D; green) structures with the apo structure of *H. methylovorum* (red; Goldberg, Yoshida and Brick, 1994). Co-enzyme domains have been overlaid and the star indicates the position of the 217 – 223 loop disordered in the apo structures.

3.3.2.2 Position of the active site arginine

Apart from the slight opening of the active site cleft, two significant differences are apparent between the ternary/binary states and the apo state. In all ternary monomers the Arg245 side chain is ordered and directed towards the substrate binding site where it forms hydrogen bonds with D-glycerate (Figure 3.22a). The same conformation is seen in the binary monomers (Figure 3.23a) where Arg245 makes ternary-like interactions with nearby waters and the OD1 of Asp86, despite there being no substrate present. In comparison, Arg245 within the human apo GRHPR monomers is largely disordered and appears to be pointing away from the active site, similar to the orientation seen in the apo bacterial GRHPR structure (Figure 3.23b). This movement is likely to be necessary, along with cleft opening, to allow product to leave and make the active site accessible to substrate. This differs from the structure of formate dehydrogenase (Lamzin *et al.*, 1994) in which the active site directionality of the equivalent arginine is retained between the apo and holo (ternary-like) structures.

3.3.2.3 Co-enzyme binding loop

The second significant difference is a lack of electron density for the 217 – 223 loop in the apo state (βD-αE loop; Figure 3.18). In the ternary/binary state this loop forms part of the coenzyme binding pocket, although (apart from Ser217) not directly interacting with but packing around the adenine moiety of the NADPH. The sequence of this loop region and that of the equivalent loop in related enzymes (where an apo structure is available) is given in Table 3.18. The Ser217 residue is at position 2 of the β-turn leading into this loop and normally makes two hydrogen bonds to the coenzyme (Figure 3.24). The lack of these interactions in the apo state may destabilise the following loop region, accounting for the observed lack of electron density and potentially opening the adenine pocket to allow more ready binding of new coenzyme. In all other D-2-hydroxyacid dehydrogenase apo structures, however, this loop region shows little movement when compared to its equivalent holo/binary state (where available) and is in a similar position to that of the human ternary/binary state. However, the distantly related hexameric L-Alanine dehydrogenase, which consists of monomers with a D-2-hydroxyacid dehydrogenase fold, contains an equivalent loop modeled with high *B* factors in the apo state and moves by up to 6Å between the NADH bound and unbound states (Baker *et al.*, 1998). Interestingly, in those structures

where minimal movement is observed the equivalent Ser217 position is occupied by a proline (Table 3.18). Proline cannot make the hydrogen bond interactions serine does, but due to its restricted configuration it is known to increase loop stability, particularly when present at position 2 of a β-turn as it is here (reviewed in Vieille and Zeikus, 2001). Thus, it is likely the presence of this proline residue in the majority of current structures is associated with stabilisation of this loop in the apo state. Investigation of the structure of a Ser217Pro mutant of human GRHPR and other related enzymes that do not contain this proline would potentially provide more information regarding the extent of this loops movement.

Enzyme	Sequence
Human GRHPR	AC**SLTPATE**GL
H. methylovorum GRHPR	NA**PSTPETR**YF
T. thermophilus GRHPR	HT**PLTPETH**RL
Formate dehydrogenase	NC**PLHPETE**HM
D-Hydroxyglutarate dehydrogenase	HV**PGIEQNT**HI
L-Alanine dehydrogenase	AV**LVPGRRA**PI

Table 3.18 Coenzyme loop sequence.

The sequence of the disordered region in the structure of Human Apo GRHPR (residues 217 – 223) is shown in bold. The equivalent sequences from related enzymes with available Apo structures are also listed. Enzymes are the same as those referenced in Table 3.14 except L-Alanine dehydrogenase from Baker *et al.*, 1998.

3.3.3 Active site density

As noted in the results (3.2.7.1), "substrate-like" density within the active site was initially modeled as the dicarboxylic acid oxalate with which the protein had been incubated prior to crystallisation. However the presence of continuous density extending beyond the oxalate O1 position and into a pocket within the active site suggested an alternative explanation was required. The identification of an approximately 5% contamination by the GRHPR substrate hydroxypyruvate within the purchased oxalate lead to this 2-oxo keto acid being successful modeled into the available density, producing logical hydrogen bonds (Figure 3.22a). Furthermore a significant steric clash would be produced between the O1 of oxalate and the side chain of Leu59, indicating why it could not be present (Figure 3.22c).

While hydroxypyruvate was initially refined into the density it seems unlikely that this GRHPR substrate would remain non-reduced during the incubation period prior to crystallisation, particularly as NADPH was present at a 10-fold excess. Moreover, inhibition by the product of hydroxypyruvate reduction, D-glycerate, has been previously noted for a bacterial GRHPR (Izumi *et al.*, 1990). Consequently D-glycerate was modeled to explain the "substrate" density. Theoretically it is possible to tell the difference between hydroxypyruvate and D-glycerate based on the geometry at the C2 position: hydroxypyruvate is trigonal, D-glycerate is tetrahedral. However, at 2.2Å the structural resolution is insufficient to definitively ascertain this geometry. Nonetheless, in the final model an abortive complex containing reduced coenzyme (NADPH) and reduced substrate (D-glycerate) was included, as this seems the most likely combination to remain trapped.

3.3.4 Ligand occupancy

Within each GRHPR dimer of the final model of the "ternary" crystal only one monomer represents a ternary complex, with both D-glycerate and NADPH present. These have been included in the model with 75% occupancy, which was found to be optimal for the refinement (Table 3.11). The other monomer within the dimer contains only NADPH (included at 50% occupancy), making it a binary complex. As previously discussed there is no significant conformational difference between the main chains of the ternary and binary complexes. Single site occupancy has previously been observed in the *L. bulgaricus* D-LDH structure. Unlike the human GRHPR, here there are significant (1.8Å for 315 Cα) differences between the structures of the closed holo (ternary-like) monomer and the open binary monomer (Razeto *et al.*, 2002). While the extension of the dimerisation loop in GRHPR may allow the opposing monomer to 'sense' substrate occupancy, it is difficult to see how this information could be transmitted and converted into an 'empty' closure. A more feasible explanation is provided by looking at the apo GRHPR structure, where a lack of any ligand still results in effectively the same "ternary-like" closed relationship between domains (particularly chains A – C). This suggests the potential importance of crystal lattice restraints in fixing the conformation with the sub-saturation concentration of the substrate contaminant resulting in only one substrate site being occupied per dimer.

The conformational similarity of the human apo GRHPR structure also suggests why no ligand occupancy could be modeled at 100%. The crystal lattice could be made up of a certain proportion of apo monomers as their ability to crystalise in a 'closed' conformation would not have required a disruption to the crystal packing but would result in a lower average presence of coenzyme and substrate. Another possibility is that, as binding of substrate brings an entropic cost (cf: B-factors), it is energetically unfeasible to have both active sites fully closed in a dimer at the same time.

3.3.5 Previous models of substrate binding

The identities of the residues involved in the catalytic activity of the D-2-hydroxyacid dehydrogenase family have long been established by mutational and chemical modification studies (Kochhar, Chuard and Hottinger, 1992; Kochhar *et al.*, 1992a; Taguchi and Ohta, 1993, 1994). These residues have subsequently been confirmed in the previous structural work (Goldberg, Yoshida and Brick, 1994; Dengler *et al.*, 1997; Razeto *et al.*, 2002) and correspond to Arg245, His293 and Glu274 in human GRHPR. Despite this the exact mode of substrate binding has been unclear. This was predominantly due to the lack of protein structures of a true ternary complex from this family. Previous models of binding have gathered support based on holo (ternary-like) structures, which contain coenzyme and an anion at the active site; such as FDH + azide (Lamzin *et al.*, 1994), D-LDH + sulphate (Razeto *et al.*, 2002) and CtBP1 + acetate/formate (Kumar *et al.*, 2002; Nardini *et al.*, 2003). Prior to the human GRHPR structure, the only ternary structure was that of D-HicDH in complex with an *in vitro* substrate (ketoisocaproate). However, the usefulness of this structure is limited as the substrate competes with a sulphate in the active site and, crucially, by having its domains in the open "apo" conformation (Dengler *et al.*, 1997).

Based on various structures two models have been proposed to describe the mode of substrate binding in the D-2-hydroxyacid dehydrogenase family, predominantly differing in their prescribed role for the single active site arginine (Figure 1.4). Both models explain the stereospecificity of this family by positioning the substrate in a manner that presents its *si* face to the nicotinamide NC4 hydrogen, rather then the *re* face as in L-LDH (Figure 3.23c). Hydride transfer to the *si* face of the substrate results in the formation of the D-isomer product.

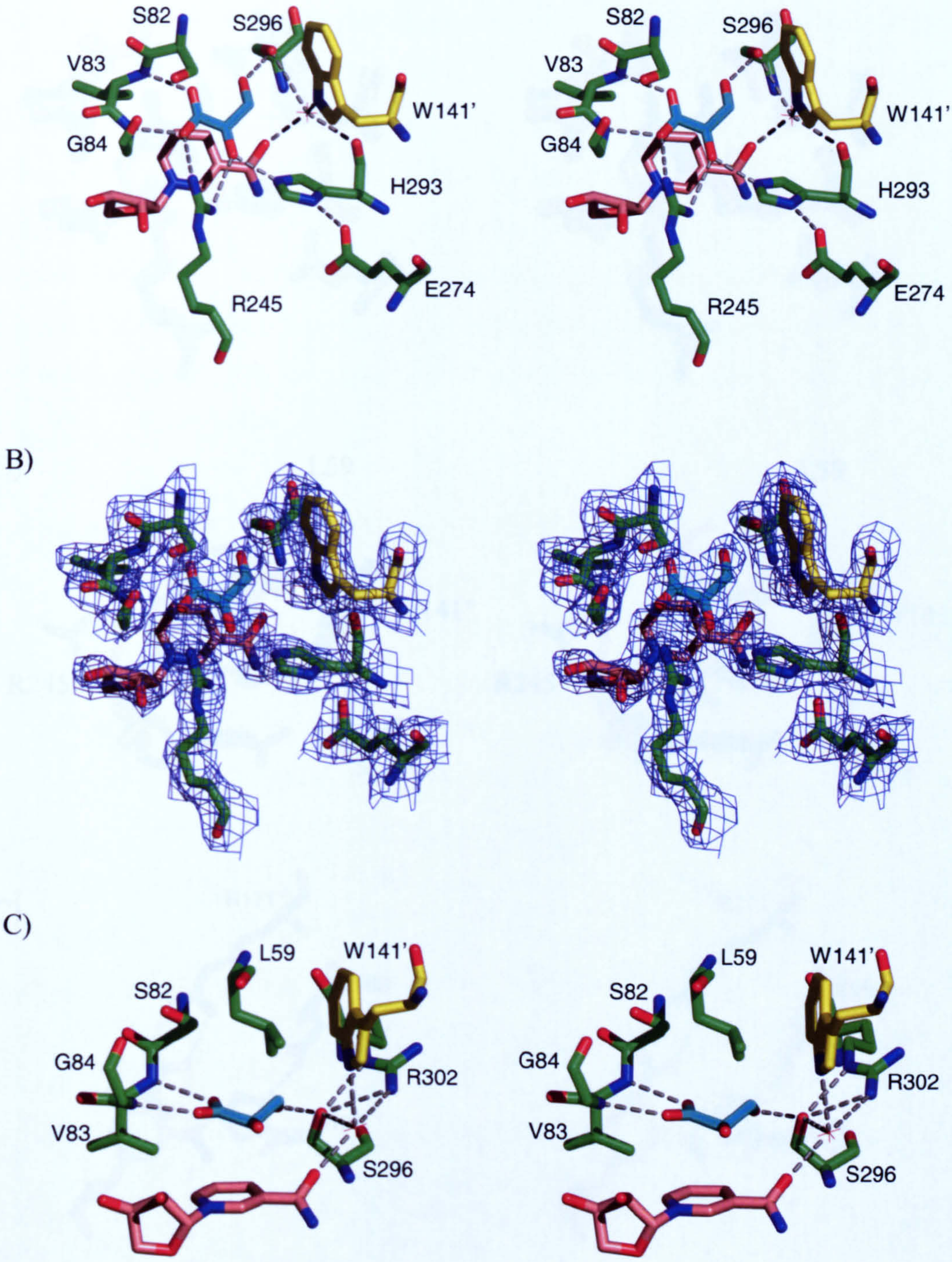


Figure 3.22 Arrangement of the active site of human GRHPR.

A) Stereoview showing D-glycerate (cyan), a fragment of NADPH (pink), and active site residues from subunit A (green) and subunit B (yellow). Hydrogen bonds are shown as dashed lines. B) As for (A), but with the addition of $2F_{obs} - F_{calc}$ electron density contoured at approximately 2 σ. C) Stereoview of active site rotated 90° in relation to (A), viewing from the position of Arg245 with residues (Arg245, Glu274 and His293) removed for clarity.

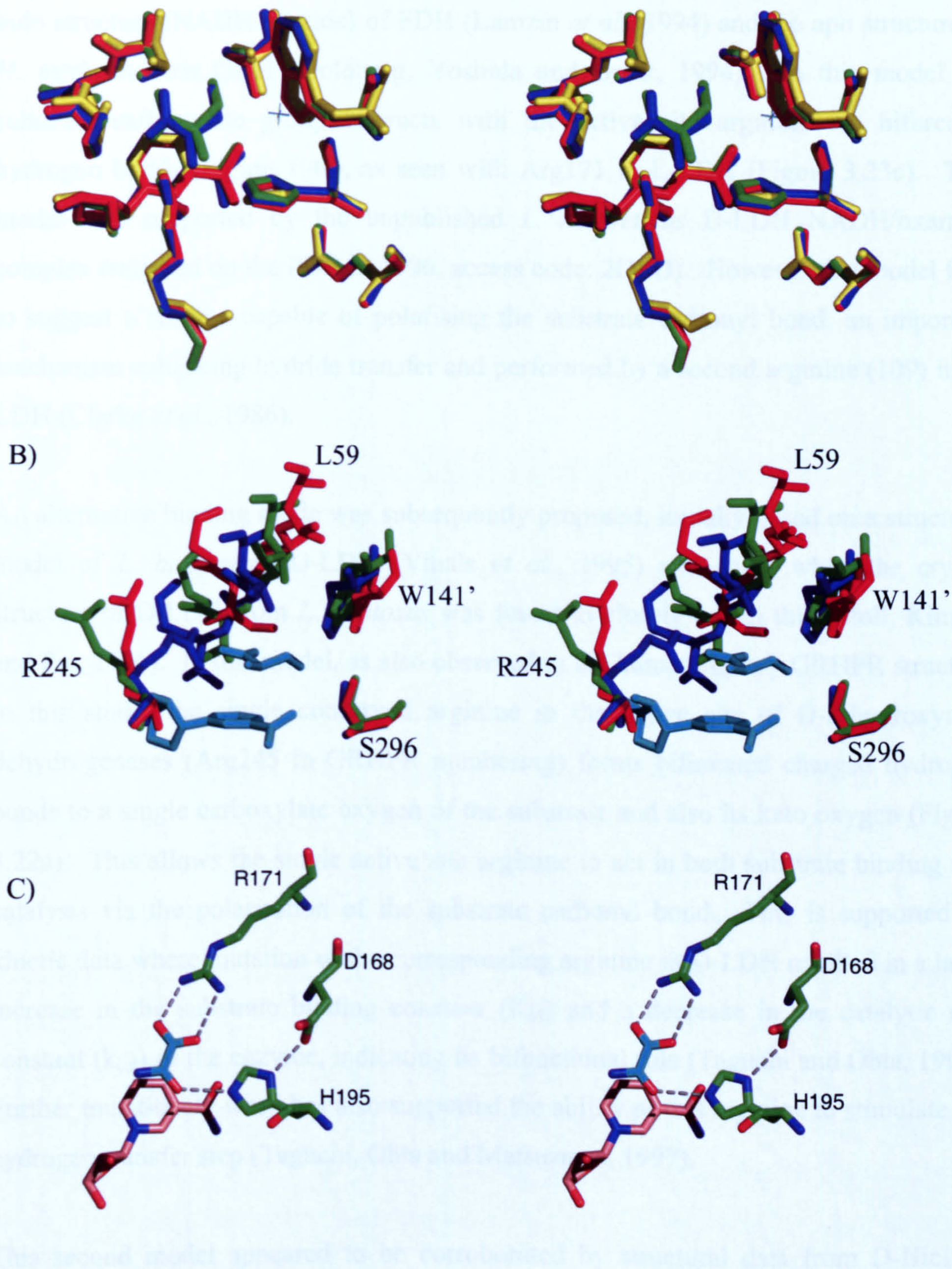


Figure 3.23 Comparisons amongst active site residues

A) Stereoview of overlay of active site residues (as in Figure 3.22) for binary and ternary GRHPR complexes: ternary subunits A (blue) and C (green), binary subunits B (red) and D (yellow). B) Stereoview of overlay of active site residues (as in Figure 3.22) for human GRHPR ternary subunit A (blue), human GRHPR apo subunit D (green) and *H. methylovorum* GRHPR apo subunit (red). D-glycerate and a fragment of NADPH are shown in cyan. C) Stereoview in similar orientation to Figure 3.22a showing the conserved active site residues from human A L-LDH (Read *et al.*, 2001).

The original model was proposed based on comparisons with L-LDH structures, the holo structure (NADH + azide) of FDH (Lamzin *et al.*, 1994) and the apo structure of *H. methylovorum* GDH (Goldberg, Yoshida and Brick, 1994). In this model the substrate carboxylate group interacts with the active site arginine via bifurcated hydrogen bonds (Figure 1.4a), as seen with Arg171 in L-LDH (Figure 3.23c). This model was supported by the unpublished *L. heleveticus* D-LDH NADH/oxamate complex (released on the PDB in 1996, access code: 2DLD). However this model fails to suggest a residue capable of polarising the substrate carbonyl bond, an important mechanism enhancing hydride transfer and performed by a second arginine (109) in L-LDH (Clarke *et al.*, 1986).

An alternative binding mode was subsequently proposed, initially based on a structural model of *L. bulgaricus* D-LDH (Vinals *et al.*, 1995) and again when the crystal structure of D-LDH from *L. pentosus* was found to closely match this (Stoll, Kimber and Pai, 1996). In this model, as also observed in the human ternary GRHPR structure in this study, the single conserved arginine in the active site of D-2-hydroxyacid dehydrogenases (Arg245 in GRHPR numbering) forms bifurcated charged hydrogen bonds to a single carboxylate oxygen of the substrate and also its keto oxygen (Figure 3.22a). This allows the single active site arginine to act in both substrate binding and catalysis via the polarisation of the substrate carbonyl bond. This is supported by kinetic data where mutation of the corresponding arginine in D-LDH resulted in a large increase in the substrate binding constant (K_M) and a decrease in the catalytic rate constant (k_{cat}) of the enzyme, indicating its bifunctional role (Taguchi and Ohta, 1994). Further mutationaly work has also supported the ability of this arginine to stimulate the hydrogen transfer step (Taguchi, Ohta and Matsuzawa, 1997).

This second model appeared to be corroborated by structural data from D-HicDH, where the *in vitro* substrate ketoisocaproate is modeled in this conformation (Dengler *et al.*, 1997). However, the sulphate ion (present at partial occupancy) competing with the substrate for the active site made this inconclusive. The structure of the transcription factor CtBP1 in complex with NAD(H) and acetate (Kumar *et al.*, 2002) or formate (Nardini *et al.*, 2003) also support this second model. However, as these are not chiral substrates and CtBP1 contains a second arginine participating in their binding, these structures were also inconclusive.

Therefore, despite the second model of substrate binding being generally accepted, the ternary structure of human GRHPR presented here is the first structure to contain an actual substrate analog bound in a 'closed' conformation, and finally confirms the orientation of substrate binding within the D-2-hydroxyacid dehydrogenase family. The position of the bound D-glycerate can be used to provide a good estimate for the binding position of related molecules. More recently, the unpublished *Aquifex aeolicus* D-LDH NAD(H)/lactate ternary complex (released on the PDB in 2007, access code: 2PI1) also replicates the substrate binding mode revealed here for human GRHPR.

3.3.6 General D-2-hydroxyacid dehydrogenase substrate binding

The fundamental interactions between GRHPR and D-glycerate are consistent with those described for the D-HicDH/ketoisocaproate complex (Dengler *et al.*, 1997) and modeled for pyruvate in D-LDH (Razeto *et al.*, 2002). The conserved active site arginine (Arg245) forms a bidentate hydrogen bond to D-glycerate, with one bond to the carboxylate oxygen (2.9Å) and the second to the keto oxygen (2.9Å, Figure 3.22a). This assists in the orientation of substrate and allows Arg245 to polarise the carbonyl bond (of the keto oxygen), lowering the energetic barrier for hydride transfer. Thus it performs the two functions that require Arg171 and Arg109 in L-LDH.

Additional hydrogen bonds are made between the D-glycerate carboxylate group and the substrate binding loop, specifically the main chain amines of Val83 (2.8Å) and Gly84 (2.8Å) further stabilising this substrate orientation (Figure 3.22). The importance in substrate binding of this loop region has been indicated by mutational studies on an adjacent residue in *L. pentosus* D-LDH. The D-LDH residue Asn97 forms a hydrogen bond to its equivalent substrate binding loop. By mutating this residue to an aspartate and consequently breaking this interaction, the affinity of the enzyme for pyruvate was drastically decreased (Shinoda, Arai and Taguchi, 2007). The equivalent residue in human GRHPR is Tyr102 and was proposed to make a similar interaction to that of Asn97 (Shinoda, Arai and Taguchi, 2007). However the Tyr102 side-chain is now seen to point away from the substrate binding loop and the equivalent bond is formed by a water mediated link to the coenzyme (NO3*, Figure 3.24). This would indicate a potential dependence of correct substrate binding on coenzyme

presence. The only change to the active site between the human and *H. methylovorum* GRHPR occurs on this loop, replacing the conserved Val83 position with an isoleucine (Goldberg, Yoshida and Brick, 1994). This seems to be the only thing that could account for the low binding constant (10.8mM) the bacterial form displays for glyoxylate (Izumi *et al.*, 1990) compared to the human form (0.24mM as will be described in Chapter 4) and *P. horikoshii* GRHPR (0.015mM, Yoshikawa *et al.*, 2007). The importance of the positioning of this loop also forms part of the discussion regarding subsequent mutational studies (Chapter 4).

The combination of Arg245 and Leu83-Gly84 position the D-glycerate *si* face (and, by analogy, substrate) parallel to the nicotinamide group of NADPH, placing the C2 of the substrate within 2.9Å of the coenzyme NC4 and hence facilitating hydride transfer (Figure 3.22). Further substrate interactions are made to the catalytic histidine (His293), which forms a 2.7Å hydrogen bond from its NE2 nitrogen to the ketone oxygen of the D-glycerate substrate (Figure 3.22). This histidine interacts with the OE1 of Glu274 which is 2.8Å distant and both stabilises and orientates the imidazole ring. The negative charge of Glu274 would be expected to stabilise the protonated state of His293, thus raising its pK_a as has been demonstrated for the equivalent residue in D-LDH (Kochhar, Chuard and Hottinger, 1992; Kochhar *et al.*, 1992a).

3.3.7 Specific GRHPR substrate interactions

Unusually, all four monomers of human GRHPR - both binary and ternary states - contain a single water molecule within the active site. It is generally considered that solvent needs to be excluded from the active site of oxidoreductases, thus lowering the dielectric environment and assisting the hydride transfer reaction (Clarke and Dafforn, 1998; Antoniou *et al.*, 2002). However, within the active site this water makes multiple hydrogen bonds (Figure 3.22): to the side chains of Trp141' (3.0Å), Arg302 (3.0Å), the carbonyl oxygen of His293 (2.9Å) and the nicotinamide moiety (2.6Å). The multiple interactions made by the water may compensate for its potentially unfavorable presence and could provide the enzyme with a level of flexibility, allowing the active site to adjust to substrates of different sizes. A similarly placed water is present in the active site of *Bacillus stearothermophilus* L-LDH and also makes multiple protein contacts (Wigley *et al.*, 1992).

The orientation of the substrate imposed by Arg245 and Leu83-Gly84 positions the C2 substituent in a pocket formed by Leu59, Leu107, His293, Ser296, Met305 and Trp141'. These residues are all well conserved across the GRHPR family (Figure 1.2) and limit the possible size of substituents on the substrate. The *H. methylovorum* GRHPR has been tested for activity towards a large number of potential keto acid substrates with only glyoxylate and hydroxypyruvate showing activity (Izumi *et al.*, 1990). From the human structure it is clear that a larger, or branched, C2 substituent would be unable to fit in this pocket. The D-glycerate (and by analogy, hydroxypyruvate) hydroxyl group makes a hydrogen bond to that of Ser296 (2.8Å) (Figure 3.22) and any non-hydrogen bonding substituent at this position would clearly cause a steric clash. Any additional length in substrate would also collide with Leu107 and Met305. Furthermore, a different substituent (or branch) at the C3 position would severely clash with Leu59, His293 and Trp141' (Figure 3.22). A key constriction occurs with the Leu59 side chain which comes within 3.2Å of the C3 carbon of D-glycerate (Figure 3.22c). This is within van der Waals distance and the unfavorable proximity is likely to be compensated for by the hydrogen bond to Ser296. It can be seen in Figure 3.23b that this active site opens slightly in the human apo form with the substrate binding loop, Arg245 and Leu59 all moving in a way that would make the site more accessible for substrate, although again this movement is more pronounced in the *H. methylovorum* structure. This opening would allow hydroxypyruvate to form a number of interactions prior to being exposed to the full Leu59 constriction which would occur upon full closure. This also shows how oxalate could have bound initially in the active site (via Arg245 and His293), thus blocking access for other substrates and acting as a competitive inhibitor. However it would prevent full closure by clashing with Leu59 and thus be rejected from the active site and consequently not present in the crystal structure despite being present in 20-fold excess relative to the hydroxypyruvate contamination. The importance of this Leu59 restriction and the compensation provided by Ser295 are ideas further explored in Chapter 4 where the catalytic activity of human GRHPR is investigated.

3.3.8 Coenzyme binding

The NADPH coenzyme is bound in the cleft between the two domains, and predominantly makes contacts with residues from the coenzyme binding domain (Figure 3.24). The elongated conformation is consistent with that seen in other dehydrogenases (Carugo and Argos, 1997) and many of the same interactions previously described for NADH in other D-2-hydroxyacid dehydrogenases are conserved (Lamzin *et al.*, 1994; Dengler *et al.*, 1997; Razeto *et al.*, 2002).

Adenine

The adenine end of the coenzyme binds in a predominantly hydrophobic pocket created by the side chains of Ile159, Thr183, Cys216, Ala221, Thr222 and Leu225 provided by βA, βB, βD and αE (Figure 3.18). A number of these residues form part of the loop (217 – 223) that is disordered in the apo state structure (as discussed in 3.3.2.3), suggesting this pocket is not preformed (at least not in the human enzyme) awaiting coenzyme but coalesces around the adenine region upon binding, burying the hydrophobic side chains. The adenine moiety forms no consistent hydrogen bonds with either the protein or conserved waters. This is in contrast to other family members which contain a direct hydrogen bond to a glutamine or asparagine, residues with no equivalent present in human GRHPR. The higher resolution (2.05Å) holo FDH structure also contains a water network linking the adenine moiety to nearby side chains (Lamzin *et al.*, 1994). A similar network could potentially be present in the human form, but is not visible at the resolution and quality of this structure. Overall, it is assumed that van der Waals contacts account for most of the interactions between the adenine and enzyme.

Adenine 2'-phosphoribose

As this is the first structure of a D-2-hydroxyacid dehydrogenase in complex with NADPH the interaction of the protein with the 2'-phosphate is of particular interest. The phosphate is situated within a pocket, surrounded on three sides by the side-chains of Arg185, Arg188 and Gln186, and below by Gly184. Members of this family with strict dependence on NADH contain an aspartate in place of this glycine (Asp175 in D-LDH, Razeto *et al.*, 2002) which would sterically clash with and electrostatically repel the phosphate moiety. Residues Arg185 and Arg188 provide a number of salt bridges to the 2'-phosphate (Arg185) and a charged hydrogen bond to the 3'-hydroxyl (Arg188,

Figure 3.24). Together they act as an "electrostatic clamp", sandwiching the 2'-phosphate. Unsurprisingly the side chains of these residues are largely disordered in the apo enzyme structure. A similar arrangement of positively charged residues is seen in other NADPH dehydrogenases, including the *E. coli* NADPH dependent shikimate dehydrogenase AroE, where the two structurally equivalent arginines (150 and 154) perform this function (Michel *et al.*, 2003), and in enzymes from the NADPH dependent short-chain dehydrogenase family (Tanaka *et al.*, 1996) where a lysine and arginine, more distant in sequence (residues 17 and 39 respectively), are used. This pocket is further discussed in 4.3.1.1 in relation to the determined coenzyme specificity of the human enzyme.

Pyrophosphate

Consistent with other structures, the pyrophosphate part of NADPH is positioned near the G-X-G-X-X-G motif (residues 160 - 165 in human GRHPR) which functions to help create an α-helix dipole at the end of αB (Figure 3.18), endowing this region with a slight positive change to neutralise the positive pyrophosphate (Wierenga, Demaeyer and Hol, 1985). A number of interactions between this region and water molecules in the ternary state are lost in the (most open) binary state subunit (chain D). These include a conserved water that tetrahedrally coordinates NO1 of the coenzyme to main-chain atoms and the link from NADPH (NO2) to the hinge region of the enzyme (Asp105 and Thr108) (Figure 3.24). The loss of these interactions potentially explains the reduced coenzyme occupancy seen in the binary state structures. The water mediated link to the hinge region is a conserved feature amongst ternary and holo D-2-hydroxyacid dehydrogenase structures (Lamzin *et al.*, 1994; Dengler *et al.*, 1997; Razeto *et al.*, 2002) and has previously been proposed to be an important contributor to the hydrogen bond network that stabilises the closed domain conformation (Razeto *et al.*, 2002).

Nicotinamide ribose

The nicotinamide ribose 2' and 3' oxygens make a number of interesting interactions. Firstly, alone amongst the current structures, NO3* interacts with the side chain of Ser217 which also binds the adenine ribose (AO4*) (Figure 3.24). This interaction, which is described in section 3.3.2.3, is likely to assist in the stabilisation of the subsequent loop. Also, both NO3* and NO2* make water mediated interactions with

regions of the substrate binding loop (Val83 and Asp86 respectively) with the Val83 interaction only being made in the ternary state monomers (Figure 3.24). These bonds are likely to be involved in stabilising the closed domain conformation and helping to correctly position this loop and, consequently, the substrate.

Nicotinamide

Hydrogen bonding interactions with Gly295, Asp269, Ile242 and the active site water result in a 33° twist (average of the 4 bound coenzymes) of the carboxamide group out of the pyridine plane. A similar degree of twist is seen in other family members. This places the carboxamide oxygen (NO7) in the *cis*-orientation relative to NC4 (Figure 3.24), consistent with other dehydrogenases, and has been proposed to activate the NC4 atom for hydride transfer (Li and Goldstein, 1992). These interactions, and those of the nicotinamide ribose, position the nicotinamide A face for hydride transfer, with 2.9Å between NC4 and the C2 position of the substrate/product, a suitable distance for transfer to occur.

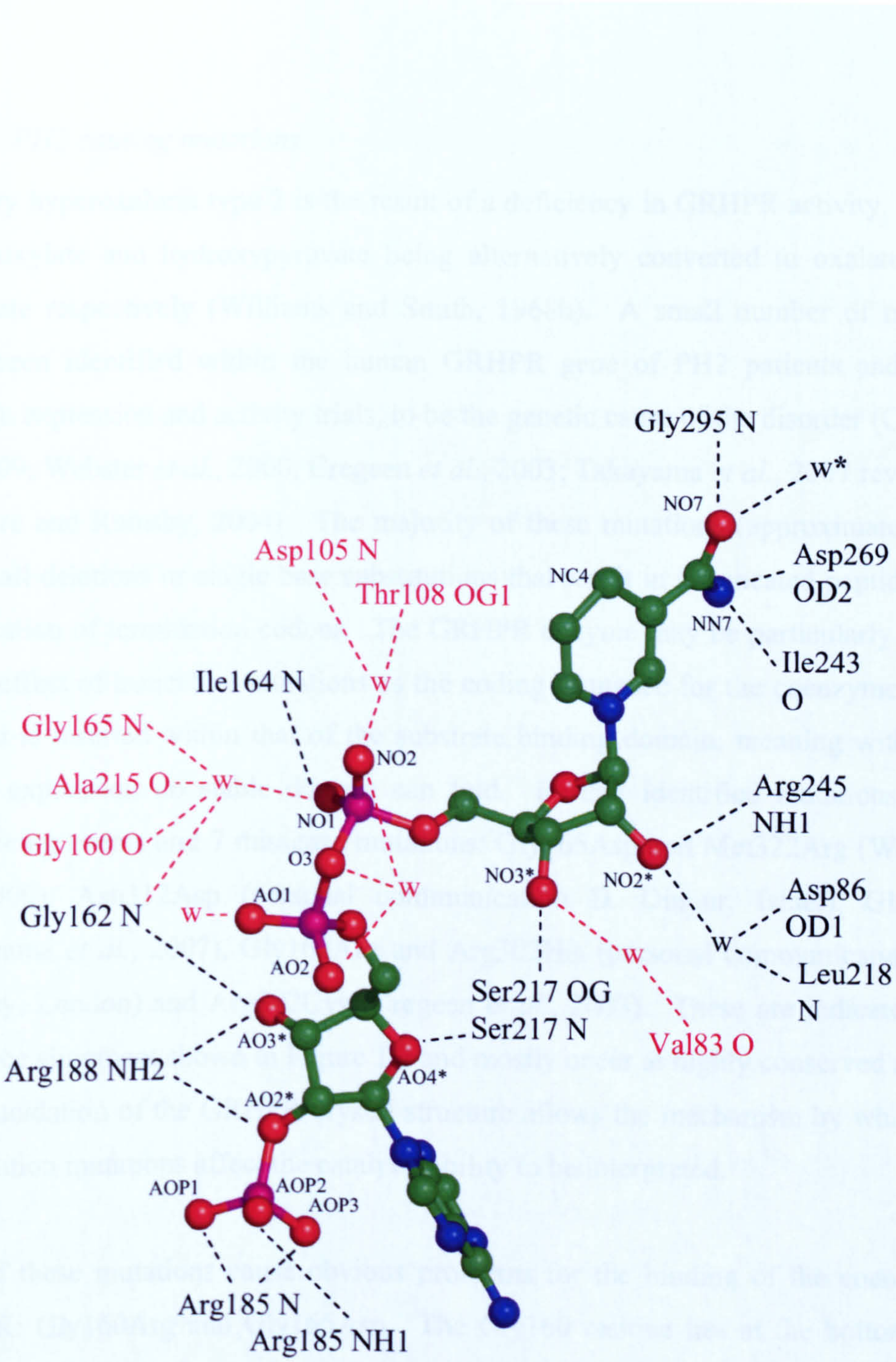


Figure 3.24 Hydrogen bonding interactions with NADPH.

The bonds formed between human GRHPR and the co-enzyme are shown as dashed lines with the atoms involved labelled. Water molecules (w) are included with their coordinating residues where appropriate. The active site water (indicated by *) makes the further interactions shown in Figure 3.22, which are not repeated here. Interactions shown in red only occur in the ternary monomers. The NADPH atom labelling is based on Carugo and Argos, 1997.

3.3.9 PH2 causing mutations

Primary hyperoxaluria type 2 is the result of a deficiency in GRHPR activity, resulting in glyoxylate and hydroxypyruvate being alternatively converted to oxalate and L-glycerate respectively (Williams and Smith, 1968b). A small number of mutations have been identified within the human GRHPR gene of PH2 patients and shown, through expression and activity trials, to be the genetic cause of the disorder (Cramer *et al.*, 1999; Webster *et al.*, 2000; Cregeen *et al.*, 2003; Takayama *et al.*, 2007 reviewed in Danpure and Rumsby, 2004). The majority of these mutations (approximately 80%) are small deletions or single base substitutions that result in a truncated peptide due to the creation of termination codons. The GRHPR enzyme may be particularly exposed to the effect of truncation mutations as the coding sequence for the coenzyme binding domain is inserted within that of the substrate binding domain, meaning without full length expression no stable domain can fold. Further identified mutations include missplicing events and 7 missense mutations: Gly165Asp and Met322Arg (Webster *et al.*, 2000), Asn312Asp (personal communication D. Dinour, Israel), Glu113Lys (Takayama *et al.*, 2007), Gly160Arg and Arg302His (personal communication Dr G. Rumsby, London) and Arg302Cys (Cregeen *et al.*, 2003). These are indicated in the sequence alignment shown in Figure 1.2 and mostly occur at highly conserved residues. The elucidation of the GRHPR crystal structure allows the mechanism by which these substitution mutations affect the catalytic ability to be interpreted.

Two of these mutations cause obvious problems for the binding of the coenzyme to GRHPR: Gly160Arg and Gly165Asp. The Gly160 residue lies at the bottom of the adenine ribose binding pocket and is highly conserved across not only the D-2-hydroxyacid dehydrogenase family but all NAD(P)H binding enzymes, being substituted only rarely for an alanine (Carugo and Argos, 1997). The inclusion of the bulky arginine side chain at this position would clearly inhibit the ability of the coenzyme to bind. This glycine is also at position 1 of the nucleotide-binding βαβ 'fingerprint', Gly-X-Gly-X-X-Gly (Wierenga and Hol, 1983), where it facilitates the required sharp turn in the polypeptide allowing subsequent interactions with the pyrophosphate region of the coenzyme. The glycine at position 165 is again highly conserved across all NAD(P)H binding enzymes (Carugo and Argos, 1997) as it constitutes the third glycine of the βαβ 'fingerprint'. Here this small residue is required to allow the close interaction between the β-strands and α-helix of this fold (Wierenga

and Hol, 1983), an interaction the mutation to aspartate would be expected to disrupt resulting in the reported loss of catalytic ability (Webster *et al.*, 2000).

Both the Met322Arg and Asn312Asp mutation result in a loss of GRHPR activity (Webster *et al.*, 2000 and personal communication D. Dinour, Israel) and investigation of the structure suggests this may result from the same mechanism: destabilisation of the C-terminus. The C-terminus of human GRHPR lies at the edge of the substrate/coenzyme binding cleft (Figure 3.17) and its destabilisation may inhibit binding and domain closure or may target the enzyme for degradation. The Met322Arg mutation may cause this by the larger arginine side chain not being accommodated in the hydrophobic pocket which accommodates the original methionine in this position, resulting in distortion of the main chain. It seems unlikely that the positive charge of the arginine would be readily accommodated in this hydrophobic region which is central to the core of the domain, and hence folding is likely to be disrupted. Similarly, substitution of Asn312 - a fairly well conserved residue – with a charged aspartate side chain also introduces a charged group close to the same hydrophobic core region, and again is likely to disrupt the domain folding.

The recently reported Glu113Lys mutation (Takayama *et al.*, 2007) occurs at the dimerisation interface. The Glu113 residue normally forms a hydrogen bond from the dimerisation loop to the coenzyme binding domain of the opposite monomer. As suggested by Takayama *et al.*, (2007) this mutation represents a change in side chain size and charge and may destabilise the dimer interface. This is likely to be particularly deleterious as Trp141 at the tip of the dimerisation loop participates in the active site of the opposing monomer.

Two PH2 causing amino acid substitutes have been identified for Arg302, one to a histidine (personal communication Dr G. Rumsby, London) and one to a cysteine (Cregeen *et al.*, 2003). As discussed in section 3.3.7, the arginine residue is in close proximity to the GRHPR active site and participates in a hydrogen bonding network involved in the stabilisation of the hydroxypyruvate substrate (via Ser296) and the dimerisation loop/Trp141' (via H_2O and Pro140') (Figure 3.22c). Neither of the substitute amino acids, while theoretically being a reasonable chemical fit, would be able to facilitate the multiple interactions that the native arginine residue performs. The

loss of these hydrogen bonds could be expected to affect dimerisation and substrate binding of GRHPR. Interestingly the Arg302Cys mutation has been shown to completely lack hydroxypyruvate reductase activity but retains some (5.6% wild type) glyoxylate reductase activity (Cregeen *et al.*, 2003). This indicates the hydrogen bonding network is of particular importance for the enzymes substrate specificity and is further discussed in Chapter 4.

These explanations for how these mutations reduce GRHPR activity hence leading to PH2 leads to the question of whether drugs could be designed to overcome this problem. The mutations that are thought to decrease protein stability (Met322Arg and Asn312Asp) or dimerisation (Glu113Lys) might potentially be rescued by pharmacological chaperones, small molecules selected to enhance protein stability. This would allow the GRHPR to regain some functionality and reduce the pathology of PH2. Research into the potential of pharmacological chaperones as a class of drugs is still at an early stage and is largely focused on cell surface and exported proteins (Loo and Clarke, 2007). This rescue mechanism has been suggested to have potential in stabilising certain mutants of AGT, the cause of the more common PH1 (Danpure, 2005). However, due to the general rareness of PH2 and the further infrequency of these 'destabilising' mutations it seems unlikely that this would become a target of specific drug design. This does not preclude alternative approaches discussed elsewhere in this thesis such as gene therapy (inducing expression of wild-type GRHPR) or the (partial) inhibition of L-LDH in order to reduce levels of oxalate, which appears to be responsible for most morbidity associated with PH2.

Chapter 4 Kinetic studies of human GRHPR

4.1 Chapter overview

The work presented in the following chapter was done in collaboration with Dr G. Rumsby (UCL Hospitals, London) and, where necessary for completeness, work carried out by her laboratory will be included and noted as such. The kinetics parameters of recombinant human lactate dehydrogenase (LDH) A, LDHB and GRHPR have been determined with a number of substrates in both the reductive and oxidative reactions. This has provided the first detailed kinetic information on human GRHPR and has allowed the relative roles of LDH and GRHPR in the removal of the reactive glyoxylate species and the production of oxalate under physiological conditions to be investigated. The mechanisms by which the 'healthy' GRHPR route is selected over LDH will be discussed. Additionally, using the structures described in Chapter 3, the specificity results presented in this chapter can be interpreted from a structural perspective. Accordingly a structure based model is proposed to explain the coenzyme and substrate specificity of this enzyme.

The kinetic analysis of recombinant GRHPR produced during this thesis and its comparison to the kinetics of human LDHA and B as discussed in the following chapter has been published (Mdluli *et al.*, 2005. The herein described model of GRPHR substrate specificity has also been published (Booth *et al.*, 2006).

4.2 Results

4.2.1 Expression and purification

In keeping with work already undertaken by the Rumsby Group (UCL Hospitals, London), protein from the unmodified pTrcHisB-HPR construct was prepared and used for the kinetic experiments. Expression from this construct proceeded identically to that described for the thrombin modified form. Within the methods chapter (2.3.11 - 2.3.14), the purification procedure is described for the production of GRHPR used for crystallisation. The procedure used to purify protein for the kinetic studies varies slightly in that the nickel column buffers used a 20mM potassium phosphate buffer pH

7.5 and 500mM NaCl and the gel filtration buffer was changed to 20mM potassium phosphate buffer pH 7.5, 100mM NaCl. This was to retain a consistency with work carried out in the Rumsby lab. Also due to the use of the pTrcHisB-HPR construct, the GRHPR retained the poly-His tag and the benzamidine column step was not carried out.

Selected fractions from the Ni^{2+} affinity step were pooled and buffered exchanged into 20mM potassium phosphate buffer pH 7.5, 100mM NaCl in preparation for purification via Superdex 75 gel filtration chromatography. The resulting A_{280} trace (not shown) consisted of a single, large, symmetrical peak that corresponds to purified GRHPR eluting at a volume of 52.7ml. The conversion equation ($\log Mr = -1.4758(V_{elution}/V_{viod}) + 6.8879$, where $V_{viod} = 41$) indicates a molecular weight for the GRHPR of approximately 98kDa, slightly bigger then the expected 80kDa tagged GRHPR dimer.

This purification procedure resulted in pure protein (Figure 4.1) which was dialysed into 20mM potassium phosphate buffer pH 7.5 overnight. Protein concentration was determined by A_{280} (MW = 40,027 Da, $\varepsilon = 29450\ M^{-1}.cm^{-1}$) and this pure protein was then used for all kinetic studies. The GRHPR used in this section retained its poly-His tag, a number of studies have shown that the presence of a terminal His-tag does not significantly affect the kinetics of an enzyme (Schmitt, Hess and Stunnenberg, 1993; Ma and Kovacs, 2000; Shoemark *et al.*, 2007). Additionally, due to the extended distance of the N-terminus from the active site of GRHPR this seemed unlikely.

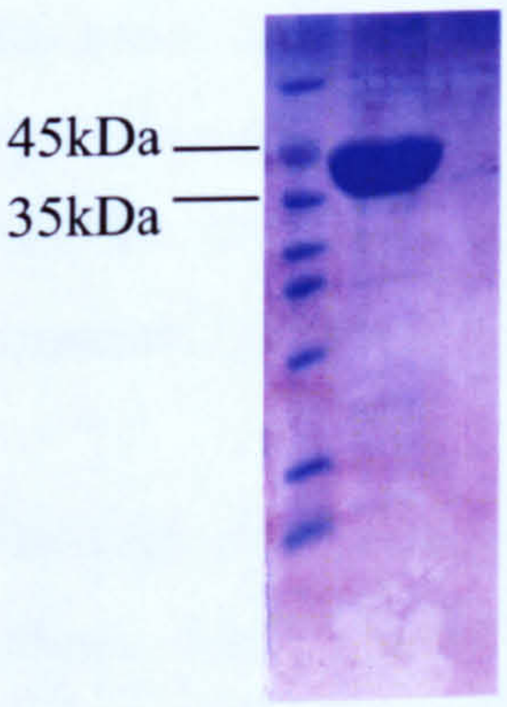


Figure 4.1 SDS-PAGE gel of GRHPR used for kinetics.

Pooled GRHPR after final purification by size exclusion chromatography showing greater the 95% purity.

4.2.2 Kinetic values to be determined

Calculation of the V_{max} and K_M parameters for each enzyme:substrate:coenzyme combination were done in the same manner, by following the decrease in absorbance as NAD(P)H was converted to NAD(P)$^+$for the reduction reactions. An increase in absorbance is seen during the oxidation reactions, as NAD^+ is converted to NADH. The values of these parameters were calculated as described in the methods (2.3.22) from the double reciprocal Lineweaver-Burk plots, where K_M is equivalent to -1/x-intercept ($1/V^{app} = 0$) and V_{max} to 1/y-intercept ($1/[S] = 0$). The K_M parameter is independent of enzyme concentration and is presented in , while the V_{max} value was used in the calculation of the rate constant as described below.

4.2.2.1 Calculation of rate and specificity constants

The catalytic rate constant (k_{cat}) is a function of the first-order rate constants for the conversion of the enzyme-substrate (ES) complex to the enzyme-product (EP) complex and subsequent dissociation (E + P), assuming initial binding steps are fast (Fersht, 1985). This value represents the maximum number of substrate molecules converted to product per active site per unit of time, providing the "turnover number" of the enzyme. The k_{cat} for the reactions were calculated using the equation $(V_{max}/60)/[E_{total}]$ and is therefore independent of the assay enzyme concentration and is reported in units of per second (s^{-1}). The k_{cat}/K_M ratio, an apparent second-order rate constant for the reaction as the rate varies with how often enzyme encounters substrate (Fersht, 1985), was also calculated. The k_{cat}/K_M value can be used as an indication of enzymatic specificity in the sense of an enzyme's preference for one or other competing substrates, ie glyoxylate or hydroxypyruvate, or one substrates 'preference' between two enzymes, ie GRHPR or LDH. The k_{cat}/K_M value are therefore also referred to as specificity constants (Fersht, 1985).

The calculated kinetic parameters are shown in and the relevant graphs are presented below. The focus of the work carried out during this thesis was on the turnover of hydroxypyruvate by recombinant GRHPR. These results shall be discussed in combination with those produced by Dr G. Rumbsy to produce a complete picture of the metabolism of substrates relevant to PH2.

4.2.3 Confirming kinetic constants for glyoxylate and NADPH

Initial work undertaken in the Rumbsy lab had varied NADPH or NADH (0.00625mM – 0.3mM) and glyoxylate (0.125mM – 1mM) concentrations, measuring the reaction rates at an enzyme concentration of 12nM (Figure 4.2). The kinetic parameters calculated from this by both curve fitting and secondary plots were the same and are given in . Also calculated were the K_M values for GRHPR with NADPH and NADH ().

In order to confirm that comparable work was carried out during this thesis, this assay was partially repeated in our laboratory. Using an excess of NADPH (0.3mM), the concentration of glyoxylate was varied (0.02mM – 0.125mM). An enzyme concentration of 12.5nM was used and the change in absorbance as NADPH was consumed was measured. A Lineweaver-Burk double reciprocal plot (Figure 4.3) was used to calculate the kinetic parameters (b). This produced very similar values to those of Dr Rumbsy (a), indicating the enzyme preparations and methodology undertaken were comparable to those of our collaborators. Additionally, it was shown that even in the presence of very high concentrations of pyruvate (up to 50mM) no GRHPR catalysed reduction could be detected ().

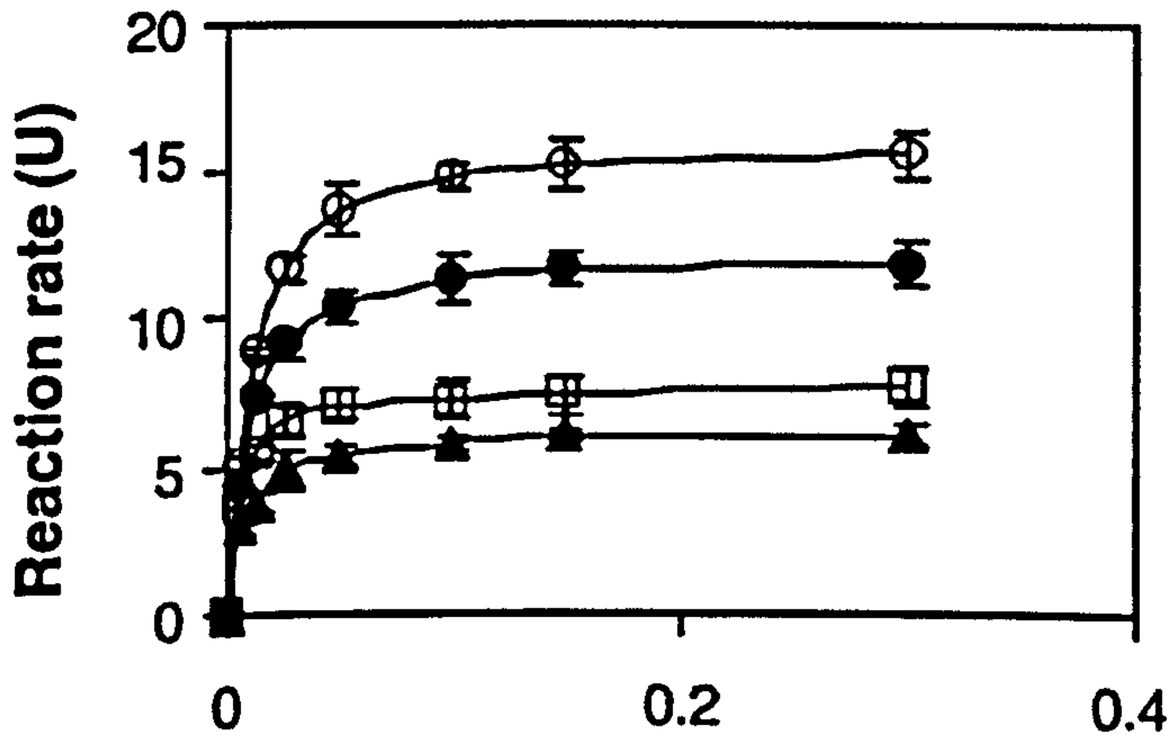


Figure 4.2 GRHPR reduction of glyoxylate with NADPH from the Rumsby lab.

Effect of variation of NADPH on reaction rate at fixed glyoxylate concentrations of 0.125mM (▲), 0.25mM (□), 0.55mM (●) and 1mM (○). Values are mean ± standard deviation, n = 3. Reproduced from Mdluli *et al.*, (2005).

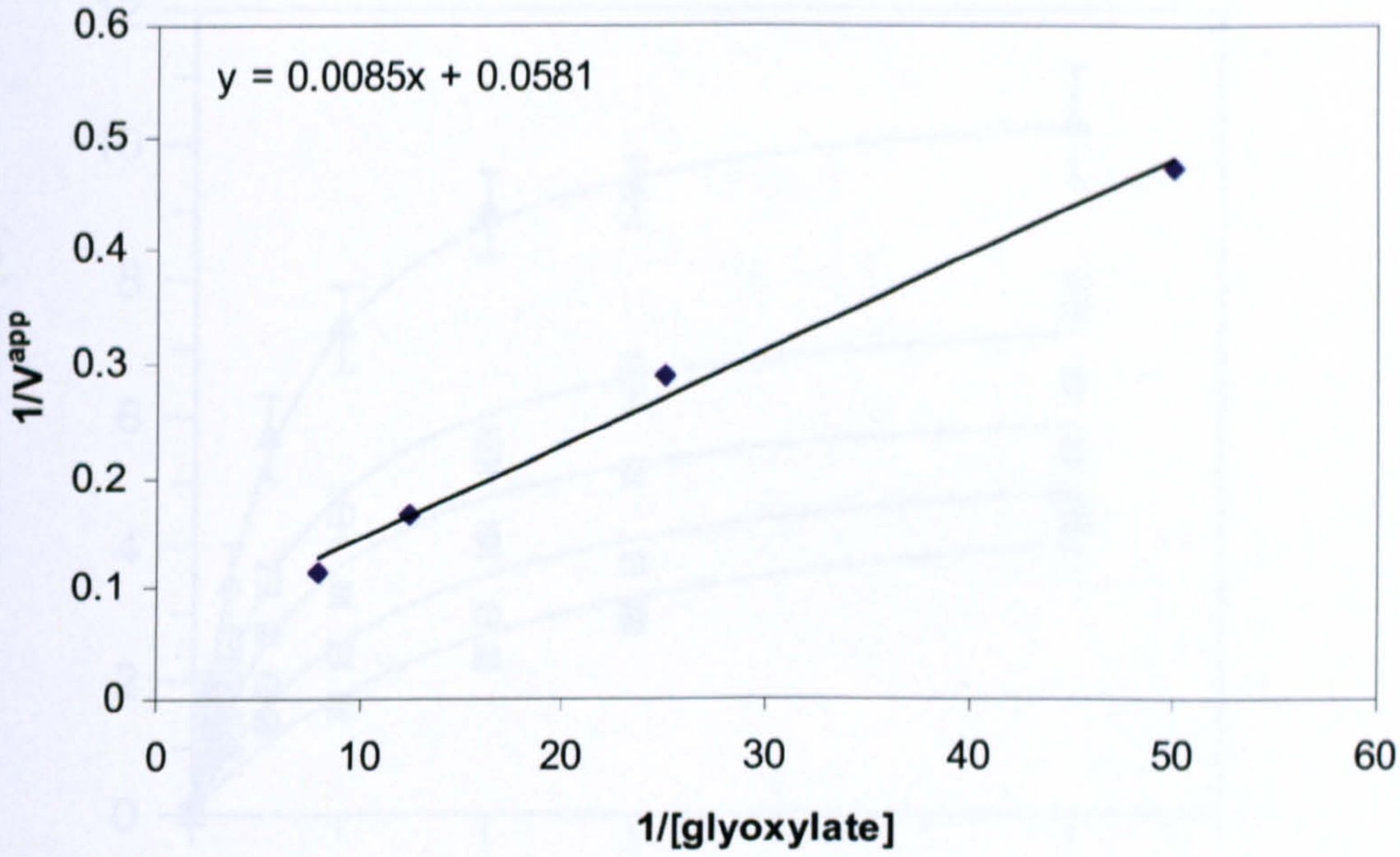


Figure 4.3 Secondary plot of GRHPR catalysed glyoxylate reduction with NADPH.

Secondary plot of $1/V^{app}$ vs. 1/glyoxylate for determination of K_M for GRHPR for glyoxylate reduction.

4.2.4 Establishing kinetic constants for hydroxypyruvate and NADPH

The kinetic parameters for GRHPR catalysed reduction of hydroxypyruvate with NADPH as the coenzyme were attempted by the Rumbsy lab in a similar manner to those for glyoxylate. The NADPH concentration was varied (0.00625mM – 0.3mM) as was the hydroxypyruvate (0.125mM – 2mM) (Figure 4.4). It became apparent that as the substrate concentration was increased the reaction rate decreased, indicative of substrate inhibition. Accurate kinetic parameters could not therefore be calculated from these data.

By methodically lowering the amount of hydroxypyruvate it was possible to observe approximately at what concentration inhibition occurred. Kinetic assays were run with 12.5nM GRHPR, excess NADPH (0.3mM) and hydroxypyruvate ranging between 0.02mM and 0.5mM (Figure 4.5) to determine the substrate concentrations best suited for study. It can be seen that the reaction rate, measuring the use of NADPH, decreases as the hydroxypyruvate concentration was increased above 0.07mM (Figure 4.5). Additionally, measurements taken at low substrate concentrations (0.02mM and 0.03mM) could not be accurately measured due to the signal to noise ratio. Data from 0.04mM to 0.07mM hydroxypyruvate was subsequently plotted on a Lineweaver-Burk double reciprocal plot (Figure 4.6) to allow the calculation of the kinetic parameters (c).

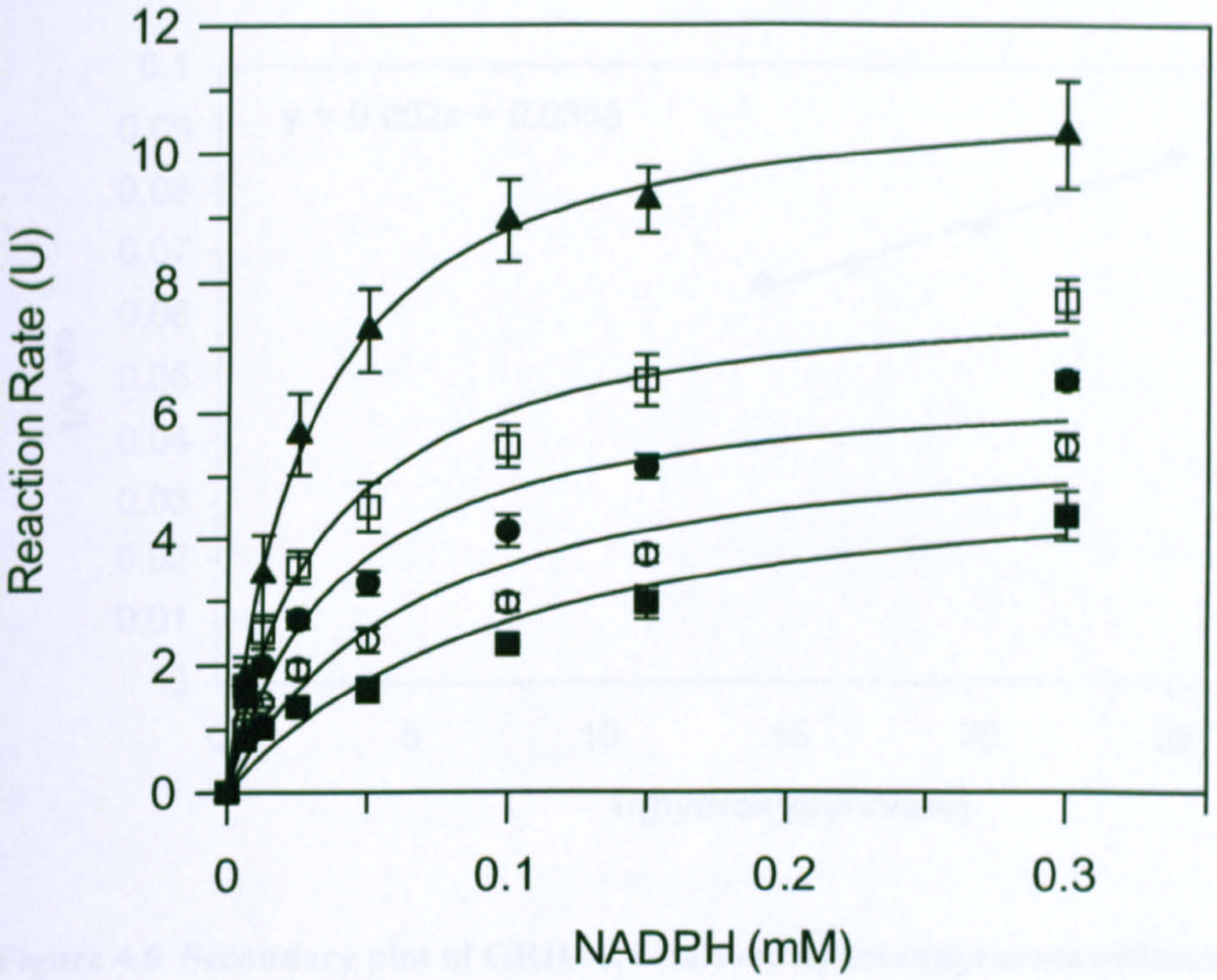


Figure 4.4 Initial plot of GRHPR catalysed hydroxypyruvate reduction with NADPH.

Effect of variation of NADPH on reaction rate at fixed hydroxypyruvate concentrations of 0.125mM (▲), 0.25mM (□), 0.55mM (●), 1mM (○) and 2mM (■). Values are mean ± standard deveiation, n = 3. Reproduced from Mdluli *et al*., (2005).

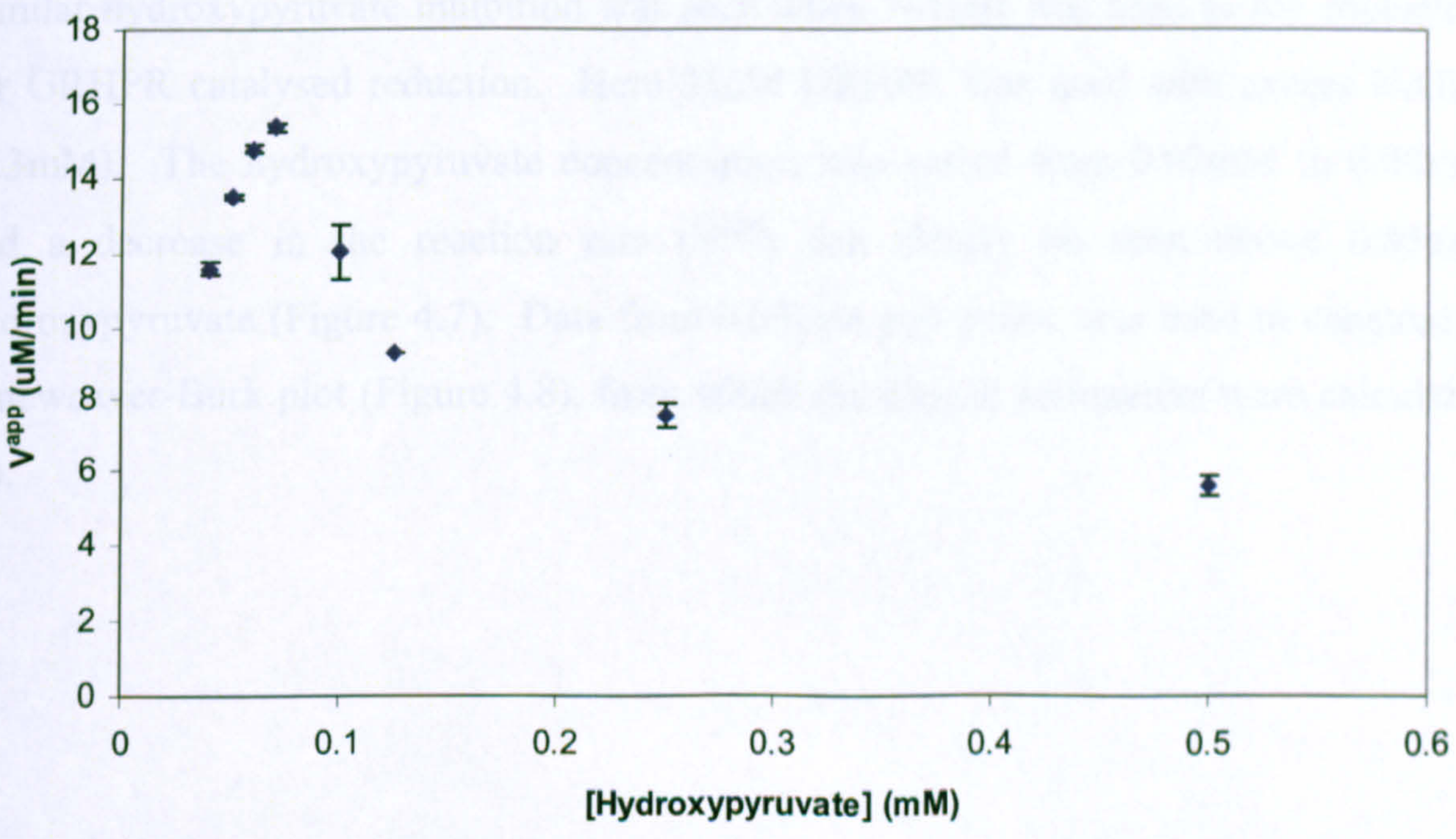


Figure 4.5 Primary plot of GRHPR catalysed hydroxypyruvate reduction with NADPH.

Effect of variation of hydroxypyruvate concentration at a fixed NADPH concentration (0.3mM). Values are mean ± standard deviation, n = 3.

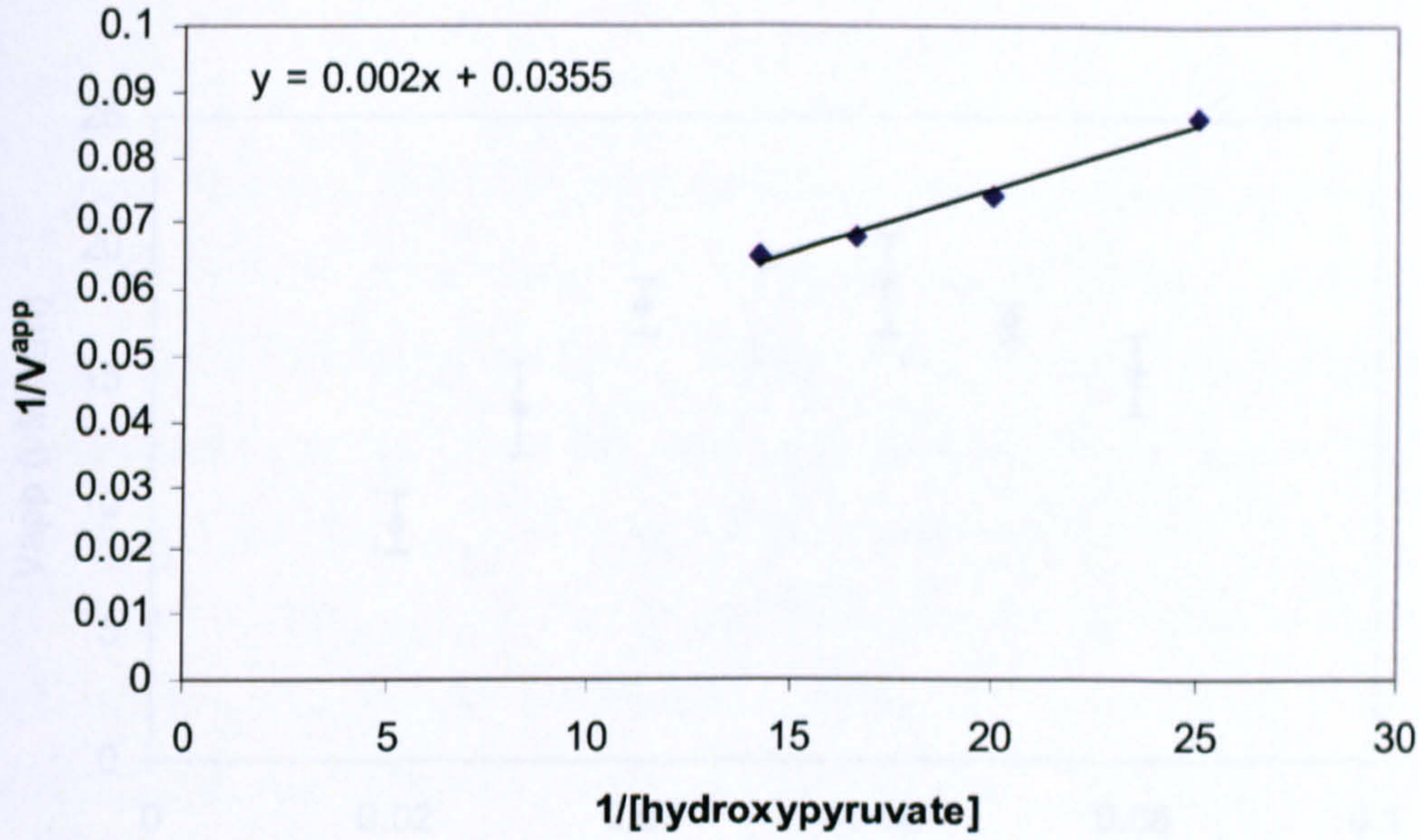


Figure 4.6 Secondary plot of GRHPR catalysed hydroxypyruvate reduction with NADPH.

Secondary plot of $1/V^{app}$ vs. 1/hydroxypyruvate for determination of K_M for GRHPR for hydroxypyruvate reduction.

4.2.5 Establishing kinetic constants for hydroxypyruvate and NADH

Similar hydroxypyruvate inhibition was seen when NADH was used as the coenzyme for GRHPR catalysed reduction. Here 25nM GRHPR was used with excess NADH (0.3mM). The hydroxypyruvate concentration was varied from 0.02mM to 0.08mM and a decrease in the reaction rate (V^{app}) can clearly be seen above 0.05mM hydroxypyruvate (Figure 4.7). Data from 0.05mM and below was used to construct a Lineweaver-Burk plot (Figure 4.8), from which the kinetic parameters were calculated (a).

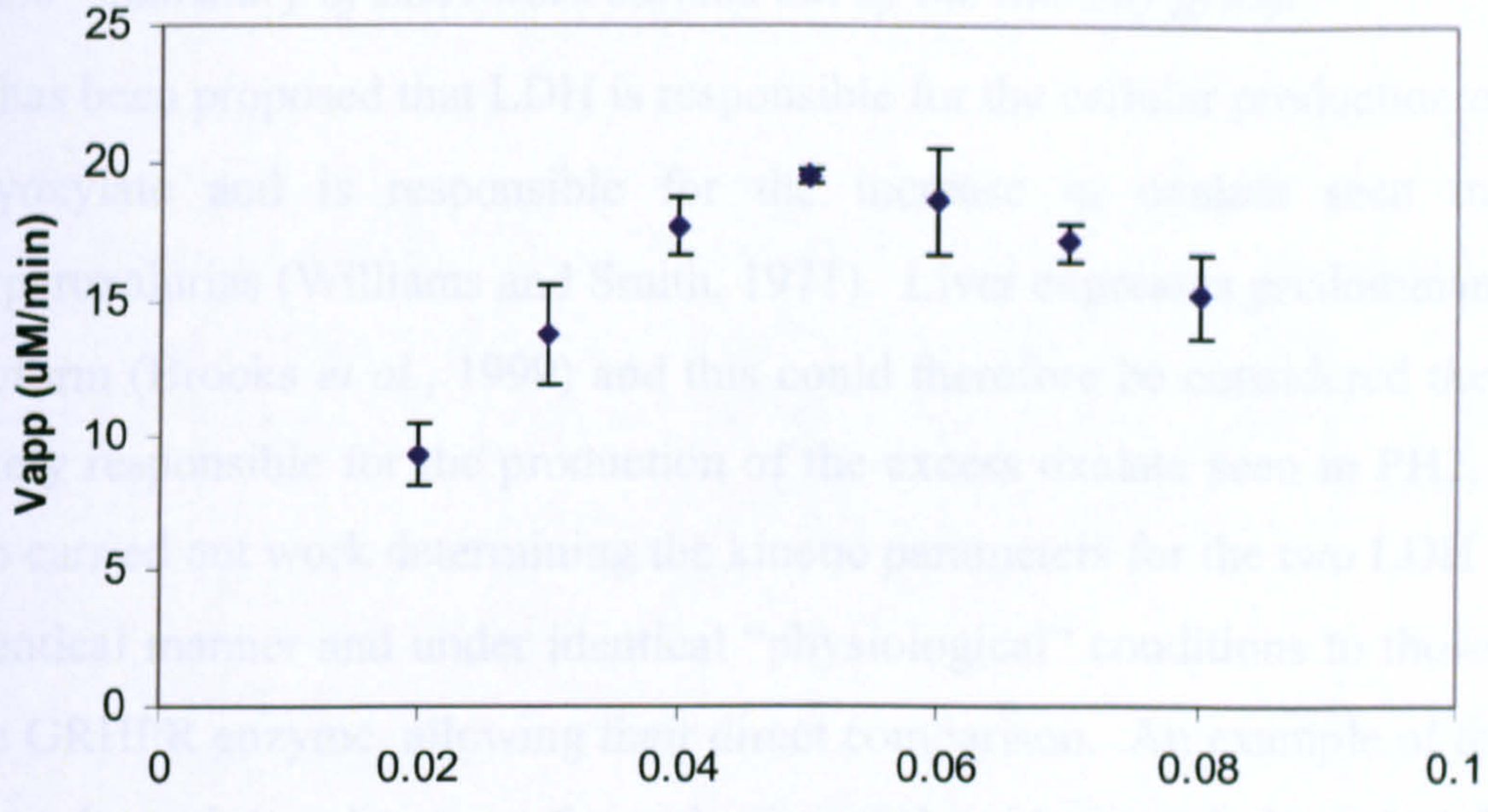


Figure 4.7 Primary plot of GRHPR catalysed hydroxypyruvate reduction with NADH

Effect of variation of hydroxypyruvate concentration at a fixed NADH concentration (0.3mM). Values are mean ± standard deviation, n = 3.

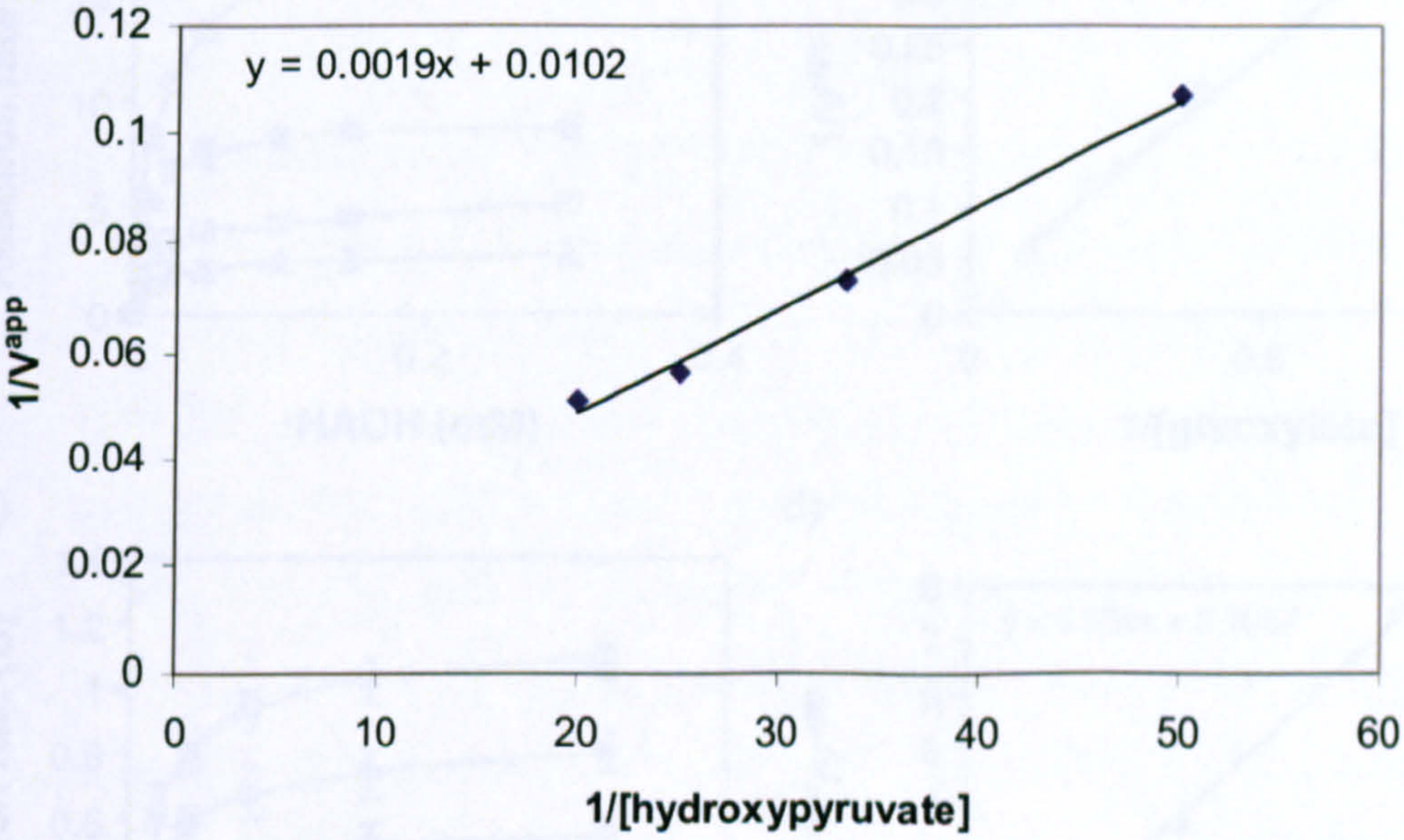


Figure 4.8 Secondary plot of GRHPR catalysed hydroxypyruvate reduction with NADH.

Secondary plot of $1/V^{app}$ vs. 1/hydroxypyruvate for determination of K_M for GRHPR for hydroxypyruvate reduction with NADH.

4.2.6 Summary of LDH work carried out by the Rumsby group

It has been proposed that LDH is responsible for the cellular production of oxalate from glyoxylate and is responsible for the increase in oxalate seen in the primary hyperoxalurias (Williams and Smith, 1971). Liver expresses predominantly the LDHA isoform (Brooks *et al.*, 1999) and this could therefore be considered the isoform most likely responsible for the production of the excess oxalate seen in PH2. The Rumbsy lab carried out work determining the kinetic parameters for the two LDH isoforms in an identical manner and under identical "physiological" conditions to those described for the GRHPR enzyme, allowing their direct comparison. An example of the primary and secondary plots relating to the reduction and oxidation of glyoxylate by LDHA are given in Figure 4.9 and the resulting kinetic parameters are presented for both isoforms in .

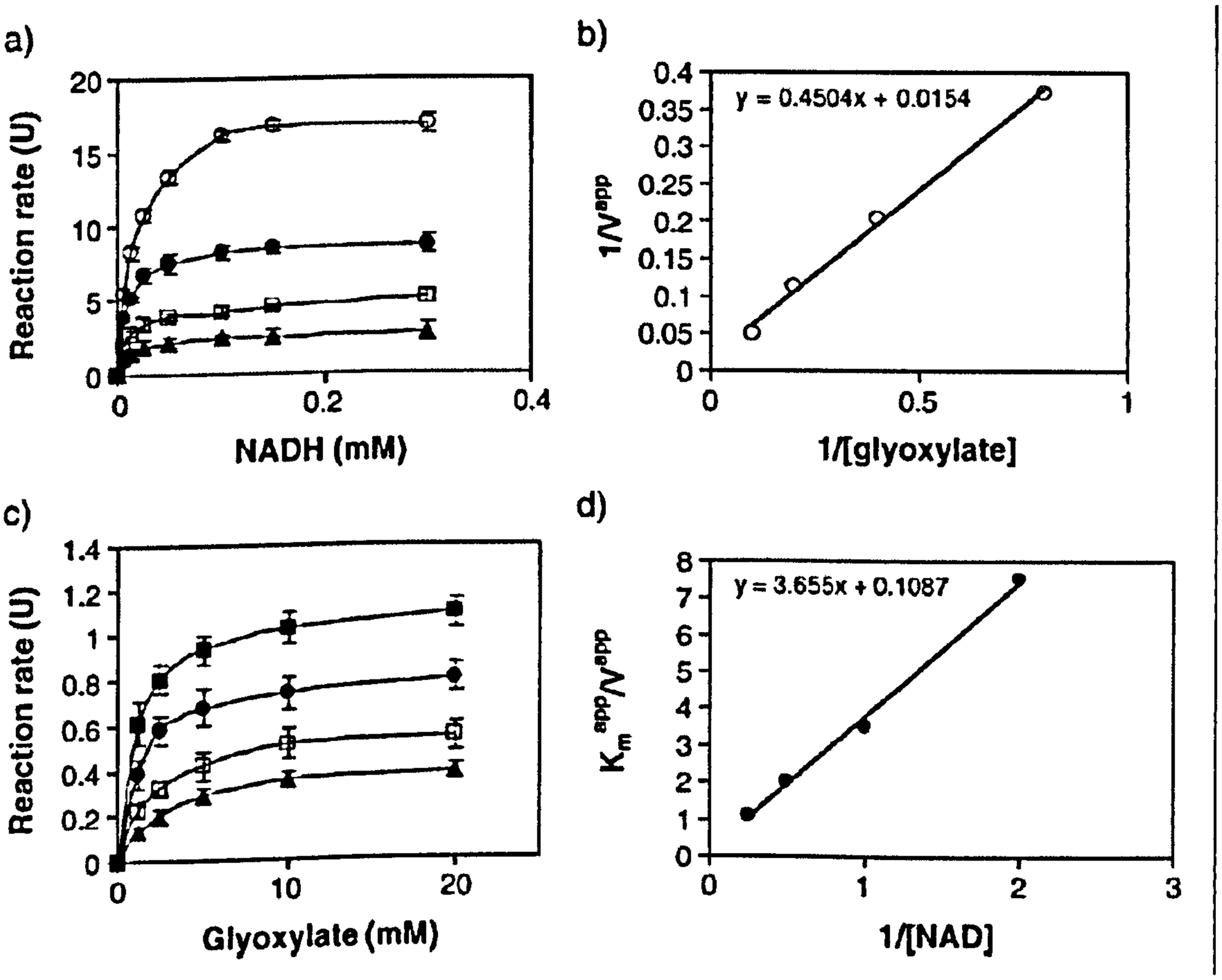


Figure 4.9 Glyoxylate reduction and oxidation by LDHA.

(Legend on following page)

Figure 4.9 Primary and relevant secondary plots for glyoxylate reduction and oxidation by human L-LDHA. Reduction of glyoxylate. (a) Effect of variation of NADH concentration on reaction rate at fixed glyoxylate concentrations of 1.25mM (▲), 2.55mM (□), 5mM (●) and 10mM (○). Values are mean ± standard deviation, n = 3. (b) K_M and V_{max} were derived from the secondary plot. Oxidation of glyoxylate. (c) Effect of variation of NAD concentration on reaction rate at fixed glyoxylate concentrations of 0.5mM (▲), 1mM (□), 2mM (●) and 4mM (■). Values are mean ± standard deviation, n = 3. For the oxidation of glyoxylate, the K_M was determined from the intercept on the ordinate of the secondary plot (d) and V from a plot of $1/V^{app}$ vs. 1/NAD (not shown). Reproduced from Mdluli *et al.*, (2005).

4.2.7 Results summary

4.2.7.1 GRHPR overview

The values produced in the Rumbsy lab for GRHPR catalysed glyoxylate reduction (with NADPH as coenzyme) shall be the values used for further discussion, although they are effectively the same as those produced in this thesis ($110s^{-1}.mM^{-1}$ against $157s^{-1}.mM^{-1}$,). This is for consistency with the publication arising from these results (Mdluli *et al.*, 2005). The reduction of both hydroxypyruvate and glyoxylate could be performed by GRHPR utilising either NADPH or NADH as the coenzyme, although the specificity constants for both substrates show a preference for NADPH, reflecting the lower substrate K_M values produced when this coenzyme is used (). Potential structural reasons for this coenzyme specificity will be discussed later (4.3.1.1). With either NADPH or NADH as the coenzyme, hydroxypyruvate (specificity constants $670s^{-1}.mM^{-1}$ and $344s^{-1}.mM^{-1}$ respectively) was the preferred substrate over glyoxylate ($110s^{-1}.mM^{-1}$ and $67s^{-1}.mM^{-1}$). Substrate inhibition of GRHPR was clearly present when hydroxypyruvate was used as the substrate, with either NADPH or NADH as the coenzyme. Inhibition occurred above 0.07mM with NADPH and 0.05mM with NADH (Figure 4.5 and Figure 4.7). Consequently kinetics parameters were calculated from data collected at hydroxypyruvate concentrations below those that had an inhibitory effect. As with the other characterised GRHPRs, from *H. methylovorum* (Izumi *et al.*, 1990), *Methylobacterium extorquens* (Chistoserdova and Lidstrom, 1991) and rat (Dawkins and Dickens, 1965), the human form shows no activity towards pyruvate.

4.2.7.2 LDH overview

Examination of the specificity constants () for the reduction reaction of the LDH isoforms, indicates a clear substrate preference order of pyruvate>hydroxypyruvate>glyoxylate. In the case of LDHA the specificity constants decrease from 2200s^{-1}.mM^{-1} for the preferred substrate pyruvate, to 430s^{-1}.mM^{-1} and 22s^{-1}.mM^{-1} for hydroxypyruvate and glyoxylate respectively. Both LDH isoforms show very similar specificity constants for oxidation reactions involving lactate (52 - 110s^{-1}.mM^{-1}) and glyoxylate (30 - 100s^{-1}.mM^{-1}) indicating no inherent substrate preference. No activity for glycolate as a substrate was evident. As expected, when NADPH was utilised as the coenzyme any activity was beyond the sensitivity of the assay used for both LDHA and LDHB indicating a very high K_M value ().

Substrate	Coenzyme	Michaelis constant (K_M) (mM)			Turnover number (k_{cat}) (s^{-1})			Specificity constant (k_{cat}/K_M) ($s^{-1}.mM^{-1}$)		
		LDHA	LDHB	GRHPR	LDHA	LDHB	GRHPR	LDHA	LDHB	GRHPR
Pyruvate	NADH	0.16±0.01	0.063±0.005		350	440		2200	7000	
[d] Hydroxypyruvate	**NADH**	0.75±0.10	0.38±0.03	**0.19±0.05** **(0.12±0.05)**	320	810	**65** **(44)**	430	2100	**344** **(367)**
Glyoxylate	NADH	29.3±1.25	9.9±0.92	1.0±0.20	650	750	67	22	75	67
[c] Hydroxypyruvate	**NADPH**	>100	>100	**0.056±0.01** **(0.057±0.01)**	<1	<1	**38** **(38)**	<1	<1	**670** **(667)**
Glyoxylate	**NADPH**	>100	>100	[a] 0.24±0.07 [b] **0.15±0.07** **(0.31±0.1)**	<1	<1	27 **23** **(40)**	<1	<1	110 **157** **(129)**
Pyruvate	**NADPH**			**>100**			**<1**			**<1**
Lactate	NAD^+	0.96±0.11	0.50±0.11		50	57		52	110	
Glyoxylate	NAD^+	0.18±0.09	0.26±0.09		18	8.0		100	30	
Glycolate	NAD^+	>100	>100		<1	<1		<1	<1	
	NADPH			0.011						
	NADH			2.42						

Table 4.1 Summary of kinetic parameter from GRHPR, LDHA and LDHB.

Table 4.1 (Legend)

Values in bold determined by M. Booth during the course of this thesis. Letters in parenthesis are directly referenced in the main results text. Values in parenthesis are the outcome of curve fitting pre-inhibition data to the Michaels-Mention equation (GraphPad Prism version 4.03, GraphPad Software Inc.). GRHPR K_M vaules for coenzyme (NADPH and NADH) were calculated with glyoxylate as the substrate. The specificity constant data is also summarised schematically in Figure 4.10. Values are ± standard deviation, n = 3.

4.3 Kinetics discussion

Prior to the work presented above, no significant kinetic characterisation of human GRHPR had been undertaken. An important factor in this work is that it was undertaken under identical, near physiological conditions for all enzymes, substrates and cofactors rather then optimised conditions to for reduction and oxidation reactions. This has allowed the preferred route of cytosolic glyoxylate and hydroxypyruvate metabolism to be established along with its likely causes. Additionally, the recently completed ternary structure of human GRHPR allows the determined specificity to be interpreted structurally.

4.3.1 Specificity as explained by the protein structure

The ternary structure of human GRHPR elucidated during this thesis, described in Chapter 3 and since published (Booth *et al.*, 2006), allows for various aspects of the above presented kinetic results to be explained from a structural perspective. While the kinetics show which substrates and coenzymes are utilised by human GRHPR, the crystallographic structure allows the mechanisms by which these specificities are determined to be proposed.

4.3.1.1 Coenzyme specificity

The kinetic data given in indicates that reduction of both glyoxylate and hydroxypyruvate catalysed by human GRHPR proceeds with a similar efficiency immaterial of which coenzyme, NADPH or NADH, is included. A slight preference for NADPH is apparent, with an approximately 2-fold increase in specificity constants when this coenzyme is used. Additionally, the Michaelis constant (K_M) values of 0.011mM and 2.42mM for NADPH and NADH respectively indicate a 200-fold tighter

binding of NADPH to human GRHPR, calculated when glyoxylate is used a the substrate. A similar NADPH preference is also seen in the rat GRHPR (van Schaftingen, Draye and van Hoof, 1989) but is in contrast to the bacterial homolog from *H. methylovorum*, which is strictly NADH dependent (Izumi *et al.*, 1990). The investigation of the human GRHPR ternary structure, the first crystal structure of a member of the D-2-hydroxyacid dehydrogenase family that utilises NADPH, indicates the potential reasons for this preference.

A survey of (then available) NAD(P)H-protein complex structures by Carugo and Argos (1997) indicated coenzyme specificity is not a product a few specific residues, but a function of the entire adenosine ribose binding pocket. However, specificity for NADH over NADPH in dehydrogenases is usually considered to be largely determined by an aspartate or glutamate residue within this pocket. This bulky, negatively charged side chain produces an electrostatic and steric repulsion effect on the 2'-phosphate of NADPH, while providing hydrogen bonds to the 2'- and 3'-hydroxyl groups of NADH. Within the NADH specific D-2-hydroxacid dehydrogenases this is provided by Asp177 in *H. methylovorum* GRHPR (Goldberg, Yoshida and Brick, 1994) and Asp175 in *L. bulgaricus* D-LDH (Bernard *et al.*, 1995). The structurally equivalent position in human and other mammalian GRHPRs is a glycine (184), resulting in no steric clash for NADPH. The single mutation of Asp175Ala in *L. bulgaricus* D-LDH has been shown to result in lower, but approximately equivalent, affinities for NADH and NADPH (Bernard *et al.*, 1995). This suggests that the lack of the aspartate residue, while important, is insufficient to determine specificity of NADPH over NADH.

As previously stated (section 3.3.8) the 2'-phosphate of NADPH in the human GRHPR structure is situated within a pocket, surrounded on 3 sides by the side-chains of Arg185 and 188 and Gln186. Residues Arg185 and 188 provide a number of hydrogen bonds to the 2'-phosphate (Arg185) and 3'-hydroxyl (Arg188) (Figure 3.24). Together they sandwich the 2'-phosphate and are likely to be responsible for the higher affinity of the NADPH coenzyme. This binding region sequence is completely conserved in the NADPH-dependent GRHPR from the rat (MGC project team, 2004).

The ability of human GRHPR to utilize NADH with an affinity (K_M) of 2.42mM is potentially due to the position of Arg188, which would be able to make hydrogen

bonds to the 2'- and 3'-hydroxyl groups of this coenzyme (Figure 3.24). Conversely, the equivalent arginine in the strictly NADPH-dependent AroE (Arg154) is not in a position to make any NADH interactions (Michel *et al.*, 2003). The importance of this coenzyme preference shall be discussed in later sections.

4.3.1.2 Substrate specificity

The ability of human GRHPR to catalysis the reduction of glyoxylate (specificity constant of $110s^{-1}.mM^{-1}$) and hydroxypyruvate ($670s^{-1}.mM^{-1}$), while remaining inactive towards pyruvate shows the human form maintains the unusual substrate selectivity seen across the GRHPR family (Dawkins and Dickens, 1965; Izumi *et al.*, 1990; Chistoserdova and Lidstrom, 1991). These 3 substrates can be seen schematically in Figure 4.10 to differ by the substituent of the C2 carbonyl group. Glyoxylate contains a small (H-), pyruvate a medium (CH_3-) and hydroxypyruvate a large ($HOCH_2$-) substituent, with GRHPR enzymatic activity only being observed with the two extremes. The ternary structure of GRHPR presented in this thesis contains a product molecule (D-glycerate) and confirms the proposed mode of substrate binding in the D-2-hydroxyacid dehydrogenase family (described in Chapter 3). The position of the bound D-glycerate can be used to provide a good estimate for the binding position of related molecules. By investigating the atomic environment surrounding the substrate binding site, and noting the conserved residues of the GRHPR family, it becomes possible to propose an explanation for the unusual substrate specificity.

Figure 3.22 in the preceding chapter displays the key residues within the substrate binding site, all of which are highly conserved across the GRHPR family (Figure 1.2). It can be seen that the CD2 methyl group of the Leu59 side chain comes within 3.2Å of the substrates "C3" position (assuming equivalent binding as D-glycerate). This produces a steric restraint on the substituent at the "C3" position due to potential van der Waals radii clashes, with the carbon atoms of two methyl groups requiring approximately 4Å separation (Mathews, van Holde and Ahern, 2000). Consequently, the inclusion of a methyl group at this position (as in pyruvate) would result in a significant clash with the leucine side chain, reflected in a lack of activity towards this substrate (K_M > 100mM). This steric restraint also explains why oxalate was not seen in the crystal structure despite being included in the initial conditions, as the oxygen at

the equivalent position would also clash. The small substituent of glyoxylate (H-) however would produce no such clash, and as a result can be bound (K_M = 0.24mM) and reduced ($110s^{-1}.mM^{-1}$) effectively.

The Leu59 side chain could be considered a "steric gate" blocking the larger pyruvate (and similar) substrates but the ability of hydroxypyruvate, with an even bigger hydroxymethyl constituent, to bind more tightly then glyoxylate (K_M = 0.058mM) suggests it acts more selectively. The tight binding of hydroxypyruvate indicates the apparent steric clash of its CH_2 group with the CD2 methyl group of Leu59 can be overcome. A likely mechanism for this is via the hydrogen bonding of the hydroxyl group of this substrate to that of the Ser296 side chain 2.8Å distant. This bond links hydroxypyruvate into a hydrogen bonding network involving Arg302, a conserved water molecule, the backbone carbonyl of the catalytic histidine (293) and the imidazole nitrogen of Trp141' (Figure 3.22). The final residue, Trp141', is provided by the dimersiation loop of the second monomer which, penetrates further into GRHPR then other D-2-hydroxyacid dehydrogenases. The high level of conservation of these residues (Leu59, Ser296, Arg302 and Trp141) in known GRHPR sequences (Figure 1.2), but not the D-2-hydroxyacid dehydrogenase family, further supports their role in substrate specificity.

A number of mutants of GRHPR were designed and constructed in efforts to test this hypothesis (section 2.3.21). However, two of these mutants aggregated during purification and could not be further utilized; the double mutant: Leu59Val/Ser296Ala and the single mutant: Ser296Ala. The Ser296Ala mutation was designed to confirm that the hydrogen bond between this residue and hydroxypyruvate is essential to the ability of GRHPR to catalyse its reduction and thereby indirectly confirm the importance of the Leu59 steric gate. It is tempting to speculate that the instability of these two mutants is due to a loss of Ser296 in its role to coordinate the active site water and correctly position Arg302 to do so (Figure 3.22). This water is involved with the dimer interface due to Trp141' and a destabilisation of its position may affect this interface and consequently the stability of the enzyme.

The two remaining mutations; Leu59Val and Leu59Ala, were both designed in an attempt to reduce the suspected Leu59:substrate steric clash and as a result provide

human GRHPR with a certain degree of activity towards pyruvate. While these mutations could be successfully purified they showed no detectable activity towards any substrate (glyoxylate, hydroxypyruvate or pyruvate) despite concentrations of up to 100mM being tested. The abolition of the activity of the Leu59Ala mutation is perhaps due to the large change is space filled by this residue allowing excess solvent molecules to remain in the substrate binding cavity upon domain closure, negatively affecting the hydride transfer reaction (Clarke and Dafforn, 1998; Antoniou *et al.*, 2002). A close examination of the active site suggests a structural mechanism by with the activity of the Leu59Val mutant has been negatively affected. If the new valine residue maintained the orientation seen for Leu59 in the native structure it would result in a clash with the Ser82 side chain of the substrate binding loop (Figure 3.22). This loop (Ser82 – Gly84) forms one end of the substrate binding site and would be critical in correctly positioning the C2 of any substrate over the C4 of the nicotinamide ring, due to the hydrogen bonds it forms to the carboxylate group tethering one end. Additional rotamers available to a valine at this position (as analysed in Coot; Emsley and Cowtan, 2004) would also result in significant steric clashes, including with the substrate binding residue Arg245.

While the mutations created in this study have not proved useful in examining the substrate specificity hypothesis, the PH2 causing mutation Arg302Cys has been enzymatically studied (Cregeen *et al.*, 2003) and provides support for the importance of the hydrogen bonding network. Like the Arg302His mutation discussed in section 3.3.9, this Arg302Cys mutation would result in a side chain which is unable to participate in the network of hydrogen bonding interactions with Ser296, the conserved water and Trp141' observed within the native enzyme (Figure 3.22). The Arg302Cys mutant has been shown to retain some ability to reduce glyoxylate (5.6% of wild type activity) but displays a complete loss of hydroxypyruvate reductase activity (Cregeen *et al.*, 2003). This would corroborate the argument that this hydrogen bond network is essential for assisting hydroxypyruvate to overcome the steric barrier presented by Leu59, but is not required specifically for glyoxylate reduction. The reduction in glyoxylate functionality in this mutant may be a consequence of a relaxation in the dimerisation due to a disruption of the Trp141' positioning (as this residue also relies on the hydrogen bonding network, Figure 3.22) or the loss of the water mediated interaction with the NO7 atom of the co-enzyme (Figure 3.22 and Figure 3.24). Like

the Ser295Ala mutation investigated here, the Arg302Cys mutation was also found to be unstable upon purification (Cregeen *et al.*, 2003) indicating the importance of the active site hydrogen bonding network to protein stability.

4.3.2 Kinetic comparisons

The enzyme L-LDH is known to be multi functional, catalysing the reduction of both hydroxypyruvate (Meister, 1952) and glyoxylate (Meister, 1952; Sawaki and Yamada, 1966), the oxidation of glyoxylate to oxalate (Sawaki, Hattori and Yamada, 1967), as well as its usual reactions with pyruvate and lactate. Consequently the role of L-LDH in the increased production of oxalate seen in the primary hyperoxalurias, and the L-glyceric aciduria of PH2 has long been hypothesised (Williams and Smith, 1968a, 1971).

The activity of L-LDH towards glyoxylate and hydroxypyruvate has been widely studied using preparations purified from various animals and tissues (Duncan and Tipton, 1969; Romano and Cerra, 1969; Warren, 1970; Gibbs and Watts, 1973; Lluis and Bozal, 1977). These studies were performed under a range of pH conditions with preparations that may have contained a mixture of L-LDH isoforms not relevant to a liver specific disease such as PH2, where LDHA is the principal isoform (Brooks *et al.*, 1999). Additionally, published work undertaken on recombinantly produced, individual human isoforms has concentrated on the pyruvate/lactate functionality of L-LDH (Eszes *et al.*, 1996; Hewitt *et al.*, 1999), with only a small amount of work including hydroxypyruvate (Gomez *et al.*, 1997). This previous work does however allow some scope for comparison.

The K_M parameters of the LDH isoforms utilising pyruvate (LDHA; 0.16mM, LDHB; 0.063mM), produced by Dr G. Rumsby and presented in , match closely previous results performed under similar buffer conditions (pH 6 – 7.5). The data regarding lactate is more difficult to compare as all previous work has been completed under high pH (8.7 - 9.2), where reduction is preferred (Nisselbaum and Bodansky, 1963; Gomez *et al.*, 1997). However, the observed pattern of weaker binding by lactate then pyruvate, represented by a higher K_M value has been reproduced.

Despite extensive previous L-LDH work the kinetic constants presented here allow, for the first time, direct comparison of the glyoxylate and hydroxypyruvate reductase ability of the individual human LDH isoforms with those of GRHPR. Importantly the parameters have all been produced under the same, near physiological pH (7.5) and temperature (37°C). While the buffer conditions may not perfectly represent a physiological environment, the 3 enzymes can now be compared and their relative roles relating to glyoxylate and hydroxypyruvate metabolism, and consequently the pathology of PH2, explored. The important reactions in regard to PH2 are the reduction of hydroxypyruvate and the oxidation and reduction of glyoxalate. It should be remembered that to fully understand the relative role of LDH and GRHPR in the liver metabolism of these substrates, an understanding of their relative expression levels would be useful. This data is not available for GRHPR and consequently this effect cannot be taken into account. At the physiological pH and temperature used in this study, the oxidation of glycolate was not observed and shall not be further discussed.

4.3.2.1 Reduction of hydroxypyruvate

Hydroxypyruvate can be reduced *in vitro* by both GRHPR and LDH, producing D-glycerate when catalysed by GRHPR or L-glycerate when utilised by L-LDH (Figure 4.10). Under normal conditions L-glycerate is not detected as a component of urine but, during the course of PH2, its production results in L-glycericaciduria, with L-LDH being implicated (Williams and Smith, 1968b).

The generation of D-glycerate by GRHPR is more favorable in the presence of NADPH (specificity constant 670s^{-1}.mM^{-1} vs. 344 s^{-1}.mM^{-1} with NADH), resulting from tighter binding of hydroxypyruvate in the company of this coenzyme. This reaction is inhibited in the presence of high levels of hydroxypyruvate, with similar levels of inhibition being seen in the presence of either coenzyme (Figure 4.5 and Figure 4.7). The GRHPR from both Rat liver (van Schaftingen, Draye and van Hoof, 1989) and *M. extorquens* (Chistoserdova and Lidstrom, 1991) also display substrate inhibition, with the predominant effect being seen with their favoured coenzyme. It is worth noting that as the physiological concentration of hydroxypyruvate (approximately 5μM, van Schaftingen, Draye and van Hoof, 1989) is far below that at which inhibition is seen,

this effect may not be applicable in the cytosol. Product (D-glycerate) inhibition has been observed with *H. methylovorum* GRHPR (Izumi *et al.*, 1990) and the proposed presence of D-glycerate in the ternary structure (Chapter 3) suggests it may potentially occur in the human enzyme. Substrate inhibition has been observed in many dehydrogenases, including both D- and L-LDHs, which react via a compulsory order mechanism where NADH binds first and NAD^+ leaves last (Denicola-Seoane and Anderson, 1990; Kochhar *et al.*, 1992a; Hewitt *et al.*, 1999). In L-LDH this inhibition is known to be the result of the formation of a covalent adduct between pyruvate and oxidised coenzyme (NAD^+) prior to its release from the enzyme (Coulson and Rabin, 1969). This adduct has not been demonstrated in GRHPR (or D-LDH) but may be responsible for the inhibition observed. The ability of substrate or product to inhibit GRHPR may provide some level of *in vivo* regulation and have been incorporated into the proposed reaction scheme given in Figure 4.11.

The K_M values for the LDH isoforms and hydroxypyruvate (LDHA: 0.75mM, LDHB: 0.38mM) are similar to those reported previously at pH 7.4 for both human isoforms (Gomez *et al.*, 1997) and chicken LDHA (Lluis and Bozal, 1977), although k_{cat} values (and consequently k_{cat}/K_M) differ, potentially due to the lower temperature (25°C) used. The specificity constants produced here show a similar level of activity between GRHPR and the dominant liver isoform LDHA, which is only active in the presence of NADH ($430s^{-1}.mM^{-1}$). Although the LDHB isoform shows an appreciably higher specificity constant ($2100s^{-1}.mM^{-1}$), it lacks a significant presence in the liver (Brooks *et al.*, 1999) and would therefore be less relevant in normal hydroxypyruvate (and glyoxylate) metabolism.

These specificity constants would suggest that there is no significant inherent preference for hydroxypyruvate between GRHPR ($670s^{-1}.mM^{-1}$) and LDHA ($430s^{-1}.mM^{-1}$), shown schematically in Figure 4.10. Therefore, as L-glycerate is not produced in healthy individuals (Williams and Smith, 1968a) cellular conditions must themselves favour reduction via GRHPR. The mechanism for this is likely to be the difference between coenzyme choice: GRHPR can exploit NADPH while LDH cannot. It is known that within the liver cytosol the NAD^+/NADH concentration ratio is approximately 1000:1, while the $NADP^+$/NADPH ratio is 1:100 (Newsholme and Leech, 1983), with the actual concentration of NADP(H) ($0.4\mu mol.g^{-1}$) being similar to

that of NAD(H) ($0.9\mu mol.g^{-1}$) (van Schaftingen, Draye and van Hoof, 1989). Accordingly, there is a higher proportion of NADPH relative to NADH within the cell, favouring reductive reactions that involve NADPH such as that of hydroxypyruvate to D-glycerate.

Hydroxypyruvate has been proposed to have a number of effects on the production of oxalate during the pathology of PH2. The reduction of hydroxypyruvate by LDH was suggested to produce a shift in the NAD^+/NADH ratio, resulting in conditions that favour the oxidation of glyoxylate by LDH (Williams and Smith, 1971). Conversely, the presence of hydroxypyruvate was shown to have an inhibitory effect on LDH catalysed glyoxylate oxidation (Warren, 1970; Raghavan and Richardson, 1983a; Raghavan *et al.*, 1997). Hydroxypyruvate has also been suggested to autoxidise to oxalate after 48-72 hours at physiological pH and additionally induce the non-enzymatic decarboxylation of glyoxylate (Raghavan and Richardson, 1983b; Raghavan *et al.*, 1997). However, hydroxypyruvate does not accumulate during the course of PH2, likely being swiftly converted to L-glycerate (Danpure and Purdue, 1995). This would imply its inhibitory effect on glyoxylate oxidation and potential for non-enzymatic conversions are not relevant *in vivo*

4.3.2.2 Utilisation of glyoxylate

The source of glyoxylate in human cells has long been unclear, but recent work suggests the majority is derived from hydroxyproline metabolism in hepatocyte mitochondria (Knight *et al.*, 2006b). Within a healthy individual this glyoxylate is predominantly reduced to glycolate by GRHPR and only a small amount of oxalate is formed, while during PH2 the lack of GRHPR results in the predominance of its oxidation to oxalate. Up until now it has been unclear how a healthy cell favors the useful reduction route over the undesired oxidation. Additionally, the ability of the human LDH isoforms to participate in glyoxylate metabolism, while long demonstrated (Gibbs and Watts, 1973), has never been studied in detail.

Reduction of glyoxylate

Both GRHPR and LDH can catalysis the reduction of glyoxylate to glycolate (Figure 4.10 and Dawkins and Dickens, 1965; Sawaki and Yamada, 1966). Under the same,

near physiological conditions used in this study, GRHPR catalyses this reduction with similar specificity constants using either NADPH or NADH as coenzyme (110/157$s^{-1}mM^{-1}$ and $67s^{-1}mM^{-1}$ respectively). This is despite the 200-fold tighter binding of NADPH (K_M 0.011mM vs. 2.42mM). The Michaelis constant seen with glyoxylate and NADPH of 0.24mM is similar to that of the Rat form (K_M 0.5mM at pH 6.0, Dawkins and Dickens, 1965). However, the affinity for glyoxylate during reduction has been shown to vary greatly across the characterised bacterial GRHPRs (K_M values of 0.015mM - 10.8mM), likely due to their divergent metabolic requirements (Izumi *et al.*, 1990; Yoshikawa *et al.*, 2007). Unlike the human form, the rat GRHPR shows substrate inhibition with glyoxylate (van Schaftingen, Draye and van Hoof, 1989).

Alternatively, the reduction of glyoxylate by the LDH isoforms proceeds only in the company of NADH, with the resulting specificity constants (22 – $75s^{-1}mM^{-1}$) again very similar to those of GRHPR catalysed reduction (67 - $110s^{-1}mM^{-1}$). This similarity is despite the tighter affinity of glyoxylate to GRHPR and is due to the greater turnover number (k_{cat}) of the LDH isoforms (~ $700s^{-1}$ vs. ~ 30-$70s^{-1}$). Consequently, as with the reduction of hydroxypyruvate, there is no inherent preference for glyoxylate reduction to occur with GRHPR rather then LDH (Figure 4.10). As reduction by either enzyme produces the same product it is difficult to comment on their actual contribution within the cell. However, once again the critical difference of coenzyme selectivity is likely to result in GRHPR being preferred. The higher proportion of NADPH compared to NADH within the cytosol favours reduction reactions that use NADPH (Newsholme and Leech, 1983). Of the two enzymes studied here only GRHPR can utilise NADPH as the coenzyme indicating, as with hydroxypyruvate reduction to D-glycerate, cellular conditions themselves are likely to favor GRHPR.

The striking difference between the specificity constants of LDH catalysed glyoxylate and pyruvate reduction is worth noting considering their structural similarity (22-75$s^{-1}mM^{-1}$ vs. 2200-7000$s^{-1}mM^{-1}$). In an aqueous solution glyoxylate is known to exist in two states, a non-hydrated form and the more common hydrated *gem*-diol form (Rendina, Hermes and Cleland, 1984). The hydrated form is considered a structural analogue of lactate while the non-hydrated form an analogue of pyruvate, with this second form the favored substrate for reduction by LDH (Duncan and Tipton, 1969; Meany and Pocker, 1991). The non-hydrated form is likely to be favoured for GRHPR

catalysed reduction also, as not only would the extra hydroxyl group be likely to create a steric clash with Leu59 (as described in section 4.3.1.2) but nucleophilic acyl addition reactions are usually preferential to the nucleophilic displacement of hydroxide (Meany and Pocker, 1991). The fraction of non-hydrated glyoxylate under similar experimental conditions to ours has been calculated to be 0.0038 (Meany and Pocker, 1991). Consequently the amount of non-hydrated glyoxylate, the reduction substrate, present will be markedly lower than the total glyoxylate concentration. This would result in actual K_M values significantly lower then those given in and as a result, specificity constants markedly higher. Applying a similar correction to that in Meany and Pocker (1991), would result in approximate specificity constants for LDH of $7800s^{-1}mM^{-1}$ and GRHPR of $14500s^{-1}mM^{-1}$ for glyoxylate reduction, similar to those of LDH catalysed pyruvate reduction. The cytosolic hydrated:non-hydrated ratio is unknown and may be different to that reported as the dehydration reaction was shown to be susceptible to various metal and buffer ions (Meany and Pocker, 1991). However, the correction to K_M values would be the same for both LDH and GRHPR, resulting in the similarity of their specificity constants remaining.

Oxidation of glyoxylate

The most important reaction regarding the pathology of PH2 is the oxidation of glyoxylate to oxalate, a reaction that GRHPR cannot catalyse but for which LDH has been implicated (Gibbs and Watts, 1973). Pig LDHB has previously been demonstrated to maximally catalyse this reaction at the non-physiological pH of 9.3 - 10 and to have either very low, or no activity at physiological pH (Sawaki, Hattori and Yamada, 1967; Warren, 1970). This is in contrast to the results of human LDH presented here at pH 7.5, where notable activity was observed. As described above, the hydrated form of glyoxylate is likely to be the favored substrate for oxidation. As this is by far the dominant form in aqueous solution, no substrate concentration correction has been considered. Additionally, the K_M values reported here of ~ 0.2mM are significantly lower then those reported at higher pH values for rabbit LDHA (3mM at pH 9.1, Duncan and Tipton, 1969) and pig LDHB (2.5mM at pH 9.4, Warren, 1970), although similar K_M variation with pH has been noted with LDH and pyruvate (Read *et al.*, 2001).

Under the "physiological" conditions used in this study, the human LDH isoforms are clearly able to catalyse the reduction of glyoxylate to oxalate (and Figure 4.10). Interestingly this oxidative ability, as measured by the specificity constants in (30 - 100$s^{-1}mM^{-1}$), is effectively identical to the reductive reaction when catalysed by either LDH (22 – 75$s^{-1}mM^{-1}$) or GRHPR (67 – 110$s^{-1}mM^{-1}$), shown schematically in Figure 4.10. This would suggest that in themselves, there is no preference for reduction over oxidation and the control is likely to rest with conditions within the cytosol. It should be noted that oxalate is known to be an inhibitor of LDH (Neilands, 1954) and may consequently limit the amount that this enzyme can produce. However, as the oxalate is being continuously excreted and sequestered as calcium oxalate it may not build up to a cytosolic concentration where its inhibitory effect becomes relevant.

Demonstrating the ability of ambient cellular conditions to enact control, a 5-fold excess of lactate over glyoxylate was shown to inhibit rabbit LDHA catalysed glyoxylate oxidation by 50% (Poore *et al.*, 1997), likely as a result of competitive binding at the active site. Under normal conditions lactate is present in hepatic tissue at levels two orders of magnitude greater then glyoxylate (Holmes *et al.*, 1995), therefore this inhibitory effect may help explain the low oxalate production of a healthy liver. Additionally, pig LDHB is known to significantly favor glyoxylate reduction over oxidation at pH 7 (Warren, 1970) and even in the presence of a 20-fold excess of NAD^+, NADH has been shown to almost completely inhibit glyoxylate oxidation by rabbit LDHA (Holmes *et al.*, 1995). While it should be remembered that cellular NAD^+ exceeds NADH by 1000-fold (Newsholme and Leech, 1983) these competition effects show that the specific conditions and metabolic state within the liver have the potential themselves to regulate LDH. Consequently, due to cellular conditions particularly the amount of lactate present it seems unlikely the LDH is responsible for oxalate production in the presence of GRHPR, only coming to the fore upon the build up of glyoxylate due to PH2 (and likely PH1). The lack of a role for LDH in healthy patients is supported by the observation that LDH subunit deficient patients have an ordinary level of oxalate excretion (Yanagawa *et al.*, 1990). The enzyme(s) responsible for this normal level of oxalate production are unknown.

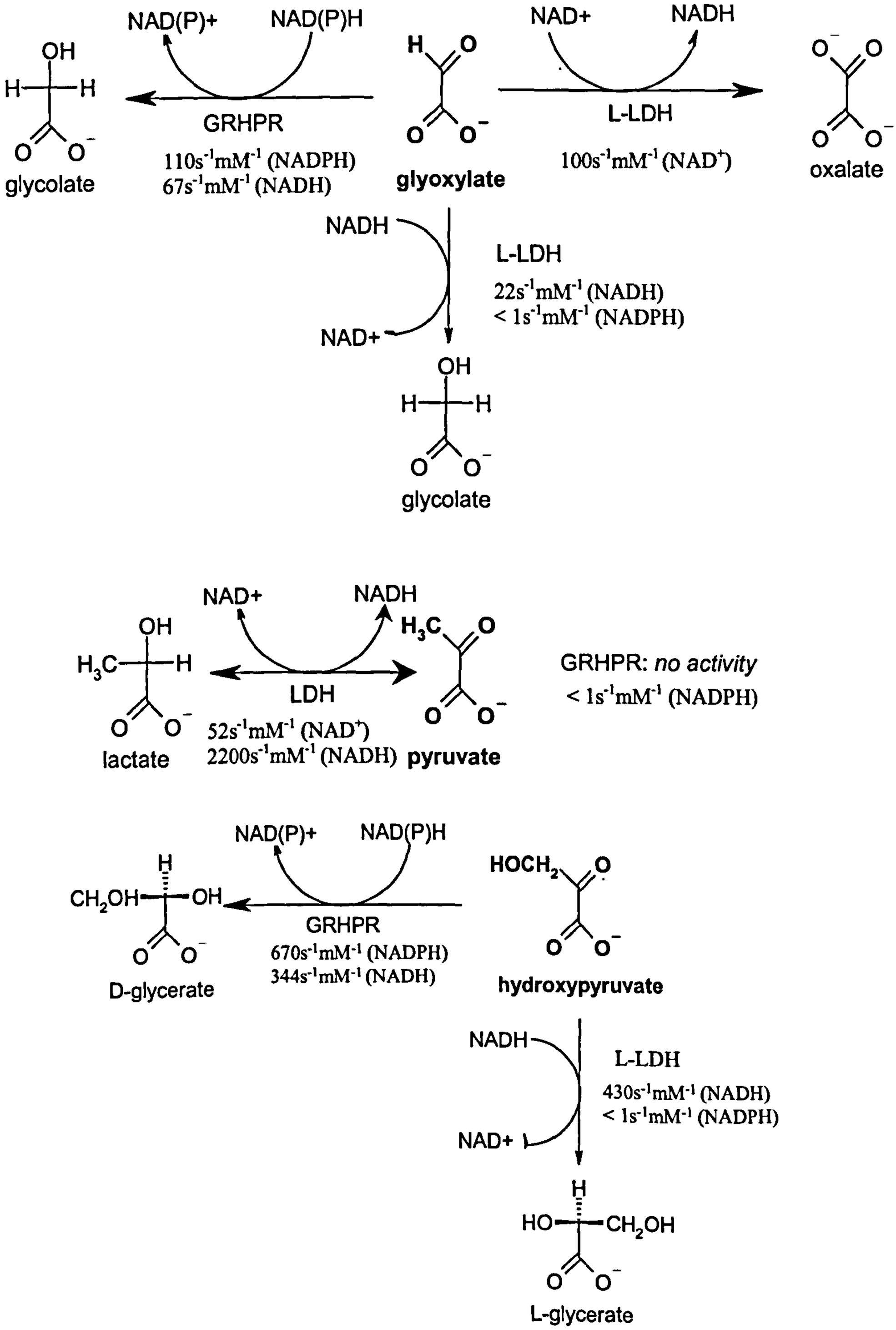


Figure 4.10 Schematic of the activity of human GRHPR and L-LDHA.

The activity of human GRHPR and L-LDHA on the related substrates glyoxylate, pyruvate and hydroxypyruvate at physiological pH. Values given are specificity constants as reported in and published in Mdluli *et al.*, (2005). Figure is adapted from Booth *et al.*, (2006).

4.3.3 Kinetic summary

Primary hyperoxaluria type 2 is caused by the loss of functional GRHPR enzyme, which results in a build up of both hydroxypyruvate and glyoxylate (Williams and Smith, 1968b). Both of these blockage points could be overcome by LDH, converting hydroxypyruvate to L-glycerate and glyoxylate to oxalate. It has therefore been proposed that LDH is responsible for producing the signature feature of PH2, L-glyceric aciduria, and the primary pathology it shares with PH1, hyperoxaluria (Figure 1.1).

In the work presented in this Chapter and published in Mdluli *et al.*, (2005) the initial kinetic parameters of recombinant human LDHA, B and GRHPR in relation to the substrates of relevance to PH2 have been presented. By investigating these enzymes under identical "physiological" conditions, it has become possible for the first time to directly compare their activities and discuss their *in vivo* relevance. A healthy person produces no L-glycerate and only a small amount of oxalate, so somewhat surprisingly the results indicate that in isolation under an approximation of physiological conditions, substrate use by GRHPR is not particularly favored over L-LDH (and Figure 4.10). The critical difference discovered was in coenzyme specificity: GRHPR can utilise NADPH whereas LDH cannot. The kinetic preference of GRHPR for NADPH was supported by the structural data where the identified 2'-phosphate binding pocket created by Arg185 and Arg188 is characteristic of NADPH-dependent enzymes. This ability to use NADPH, the more prevalent cytoplasmic reductive coenzyme, is likely to be the decisive factor that leads to the dominance of GRHPR catalysed reduction over that of LDH within the cell towards both glyoxylate and hydroxypyruvate. The lack of a physiological role for glyoxylate oxidation by LDH is probably due to the cellular predominance of its other substrates, particularly lactate. As this work shows, LDH is capable of oxidizing glyoxylate to oxalate at physiological pH. It is this reaction that causes the hyperoxaluria pathology seen in PH2 and while it is tempting to consider the moderation of L-LDH activity in order to reduce levels of oxalate during PH2, this would require a complete understanding of the many metabolic pathways to which L-LDH contributes. The substrate inhibition of GRHPR seen with hydroxypyruvate and the potential for product inhibition as displayed in the crystal structure, show how this enzyme can also, like L-LDH, be regulated by ambient concentration of its substrates and products.

A hypothetical reaction scheme for GRHPR is presented in Figure 4.11, and is based on a similar proposed scheme for D- and L-LDH (Denicola-Seoane and Anderson, 1990; Hewitt *et al.*, 1999). More work needs to be done to fully kinetically characterise human GRHPR and clarify this scheme, such as determining potential inhibition constants for substrates and products. However the work presented in this thesis has helped in the understanding of hydroxypyruvate/glyoxylate flux through the liver.

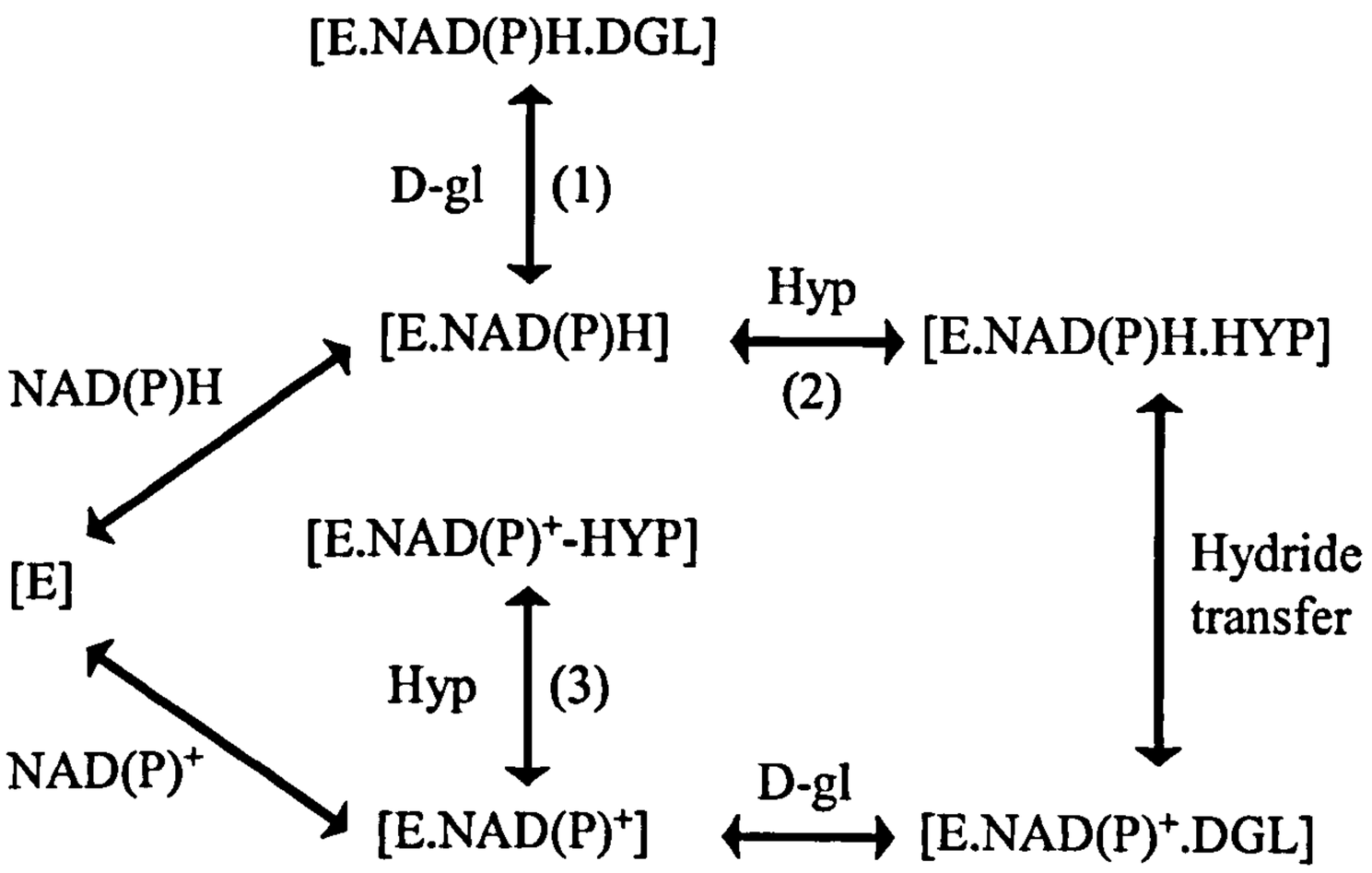


Figure 4.11 Proposed reaction scheme for GRHPR.

(1) product inhibition, as seen in crystal structure (Chapter 3), (2) binding of substrate induces domain closure (comparison between apo and ternary structures; Chapter 3), (3) potential mode of substrate inhibition, by analog with L-LDH. Hyp = hydroxypyruvate, D-gl = D-glycerate. Adapted from the reaction scheme for L-LDH in Hewitt *et al.*, (1999).

Chapter 5 Sorting Nexin 1

5.1 Introduction

The protein sorting nexin 1 (SNX1) is a member of the large sorting nexin family of proteins identified in mammals and yeast, and which have roles in cellular processes including endosomal sorting and signaling and endocytosis (reviewed in Worby and Dixon, 2002; Carlton *et al.*, 2005a; Seet and Hong, 2006; Cullen, 2008). Mammalian SNX1 is the founding member of this family (Kurten, Cadena and Gill, 1996) and the most widely studied due to its likely role in the mammalian retromer and subsequent endosomal sorting (Haft *et al.*, 2000). No crystal structures of either the whole SNX1 protein or parts thereof are available, limiting the structural knowledge of this important protein to a single NMR structure of the phox-homology (PX) domain (Zhong *et al.*, 2005). Consequently work was initiated during the course of this thesis with the aim of establishing soluble expression and purification procedures to facilitate the production of SNX1 and, where necessary, truncated forms in quantities sufficient and suitable for crystallographic studies.

5.1.1 Endosomal sorting

The endosomal system within eukaryotic cells acts to sort and deliver internalised proteins, such as various nutrient and signalling receptors, to their correct destinations (Figure 5.1). After internalisation at the plasma membrane, often via clathrin-coated vesicles, these proteins are delivered to the sorting endosome from where their trafficking decisions are made (Lemmon and Traub, 2000). Here, the transferrin receptor and many other nutrient receptors release their ligands and are recycled back to the plasma membrane. Internalised signalling receptors such as epidermal growth factor receptor (EGFR) follow the degradative path, as the final mechanism to switch off signalling. These receptors are retained within the endosome as it matures to a "late endosome", which subsequently fuses with the lysosome resulting in degradation. A further trafficking pathway is that between the early endosome and the *trans*-Golgi network (TGN). The mannose 6-phosphate receptor (MPR) is recycled between the TGN (where it binds lysosomal hydrolases) and the early endosome (where it releases

them) allowing multiple rounds of delivery. The retrieval of the MPR to the TGN appears to be directed by the mammalian retromer (Arighi *et al.*, 2004).

The exact mechanism by which different trafficking decisions are made, and the identity, function and structure of many of the proteins and protein complexes involved, have only recently begun to be elucidated. Key components of many of these trafficking pathways are members of the diverse sorting nexin family of proteins (reviewed in Worby and Dixon, 2002; Cullen, 2008).

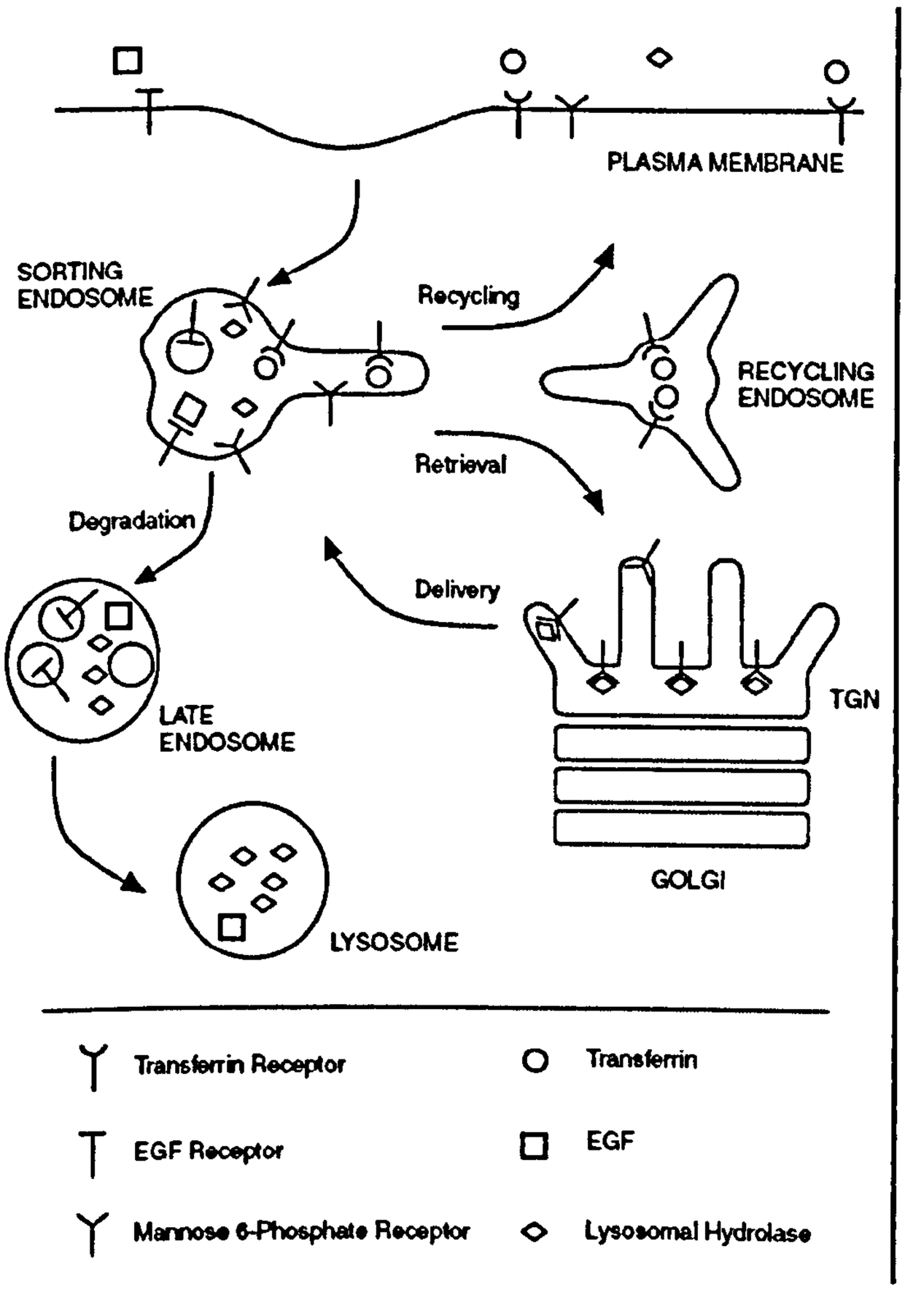


Figure 5.1 Diagram of endosomal sorting in mammalian cells.

Representation of the possible trafficking pathways undertaken by various components. Figure from Carlton *et al.*, (2005a).

5.1.2 The sorting nexin family and their important domains

Currently 33 members of the mammalian sorting nexin family and 10 in yeast have been identified (Cullen, 2008). However the functions of many of these proteins are unknown, despite many being implicated in endosomal sorting pathways. From a structural perspective the crystallographic information available is limited to the yeast homolog of SNX3 (Grd19p) which consists solely of a phox-homology (PX) domain (Zhou *et al.*, 2003) and, more recently, a truncated form of SNX9 which is involved in plasma membrane remodeling events (Pylypenko *et al.*, 2007). Additionally, only two NMR structures of PX domains from the sorting nexins have been reported (Zhong *et al.*, 2005; Song *et al.*, 2007) but provide little information on critical features such as ligand specificity and the spatial arrangements between domains (as discussed below).

The sorting nexin family of proteins is grouped together on the basis of inclusion of a particular subgroup of the PX domain superfamily, now known as SNX-PX (Teasdale *et al.*, 2001). The PX domain is a binding domain for the phosphoinositide head group of certain phospholipids (phosphatidylinositol phosphates; PtdInsP) and as such plays an important role in localising proteins to specific membranes. Many SNX-PX domains, including that in SNX1, have an affinity for PtdIns(3)P (Cozier *et al.*, 2002), which is enriched on sorting endosomal membranes (Gillooly *et al.*, 2000). SNX1-PX has also been shown to bind PtdIns(3,5)P_2 (Cozier *et al.*, 2002) which is implicated in protein trafficking, particularly involving the endosome to lysosome transition and is enriched on late endosomes (Gillooly *et al.*, 2000). The specificity of SNX9 for PtdIns(4,5)P_2, which is enriched (under certain stimuli) at the plasma membrane, is consistent with its role in clathrin-mediated endocytosis (Pylypenko *et al.*, 2007). The presence of these membrane specific domains in all sorting nexins suggests a role for all of these proteins in membrane trafficking events. It is not yet possible to predict phosphoinositide specificity of a PX domain from sequence data alone, but this might be facilitated through knowledge of the three-dimensional structure of these domains. This would permit the assignment of potential localisations for poorly studied sorting nexins, assisting with the study of this entire protein family.

While the presence of a SNX-PX domain is the defining feature of this family, a variety of different lipid binding and protein interacting domains are also present and often determine the various roles for separate sorting nexins. For the current work the most important is the Bin/Amphiphysis/Rvs (Bar) domain, but other domains including SH3 (Src-homology-3) and RGS (regulators of G-protein signalling) domains are present on various sorting nexins, suggesting (poorly understood) potential roles in integrating sorting and signalling (Cullen, 2008).

A large subset of the mammalian sorting nexins (12 of the 33) combine the PX domain with a C-terminal Bar domain. This subset includes those sorting nexins implicated in the retromer (SNX1, -2, -5 and -6, section 5.1.3) and those closely related to SNX9 and concerned with endocytosis (SNX18 and -30, Pylypenko *et al.*, 2007). The Bar domain represents a dimerisation motif which produces a curved binding surface with which the associated protein can bind membranes of high curvature, or in some cases, induce the production of curved membranes (Carlton *et al.*, 2004; Peter *et al.*, 2004 and reviewed in Habermann, 2004; Gallop and McMahon, 2005). The curvature of this dimerised domain has been seen to vary amongst the published Bar structures, with the predicted membrane diameter that it would fit varying from 150 Å in arfaptin 2 (Tarricone *et al.*, 2001) to 600 Å in PHC-F-Bar (Shimada *et al.*, 2007), with many expected to bind a diameter of around 220Å (Peter *et al.*, 2004; Weissenhorn, 2005; Casal *et al.*, 2006). The recent structure of SNX9 included its PX and Bar domains and is to date the only sorting nexin Bar domain structure known. In this structure a curvature equivalent to a membrane diameter of 400Å was seen (Pylypenko *et al.*, 2007). The sequence homology between SNX1 and SNX9 is only around 14% and therefore it is unclear whether these dimensions also apply to SNX1. More critically, while the SNX9 Bar domain has been seen to be largely insensitive to the curvature of liposomes, an indication taken to suggest its efficient tubulating ability (Pylypenko *et al.*, 2007), the SNX1 Bar domain is highly sensitive, only binding liposomes below a certain size (Carlton, 2004). The SNX1 Bar domain appears novel in its high sensitivity to curvature and yet it retains some ability to tubulate membranes (Carlton *et al.*, 2004) whereas the closely related SNX2 Bar domain can only sense but not induce curvature (Carlton *et al.*, 2005b). The exact mechanism and structural requirements by which these domains cause tubulation is unclear, with some Bar domains requiring a N-terminal amphipathic helix (N-Bar, McMahon and Gallop, 2005) which the Bar

domains of the sorting nexins are not predicted to contain. Consequently this will only fully be understood when a variety of Bar domain structures from the sorting nexins are known.

Other then the role of SNX9 in clathrin-dependent endocytosis (Soulet *et al.*, 2005) and SNX1 in the retromer (Griffin, Trejo and Magnuson, 2005) little is know about the functions of the many other members of this family. Consistent with the central importance of endosomal sorting and signalling pathways, evidence is growing for a role for SNX proteins in a range of diseases, including in the trafficking of prion protein (Heiseke *et al.*, 2008) and the amyloid precursor protein (Lee *et al.*, 2008). However, sorting nexins have been most comprehensively studied as the targeting component within the retromer, where both the PX and Bar domain are essential.

5.1.3 Retromer

The retromer is a large, multimeric complex identified in yeast to be essential for the endosome-to-Golgi retrieval of vacuolar protein sorting receptor Vps10p, the functional homolog of the mammalian MPR (Seaman, McCaffery and Emr, 1998). The retromer is thought to act as a vesicle coat complex, and has been shown to interact directly with the endosomal membrane. In yeast, this complex is formed from two subcomplexes: Vps35p, Vps29p, and Vps26p, the cargo selection complex and Vps5p and Vps17p, as a complex with a structural role (Seaman *et al.*, 1997; Seaman, McCaffery and Emr, 1998). Human orthologs of all components have been identified, and h(uman)Vps35p, hVps29p, and hVps26p shown to form multimeric complexes (Haft *et al.*, 2000). The human functional orthologs of Vps5p and Vps17 are all members of the sorting nexin family, with SNX1 and SNX2 corresponding to Vps5p (Griffin, Trejo and Magnuson, 2005) and SNX5 and SNX6 to Vps17 (Liu *et al.*, 2006; Wassmer *et al.*, 2007), although the exact sorting nexin composition within the retromer is unclear. Yeast-2-hybrid studies have also shown the expected interaction between hVps35p (the putative 'core' of the retromer) and SNX1 (Haft *et al.*, 2000). The SNX dimer is considered the targeting/structural component of the retromer and is required to recruit the rest of the retromer to the endosome. This interaction has been shown to occur between Vps35p and the N-terminus of Vps5p (Seaman and Williams, 2002). The same region is

involved between SNX1 and hVps35p (Gullapalli *et al.*, 2004) and corresponds to the largely unstructured region in SNX1 (Figure 5.3).

The structures of two retromer components, hVps29p (Collins *et al.*, 2005; Wang *et al.*, 2005) and hVps26p (Shi *et al.*, 2006; Collins *et al.*, 2008) have been solved individually. The crystal structure of a truncated hVps35p:hVps29p complex (Hierro *et al.*, 2007) provides further insights into retromer formation and lead to the current model of the retromer (Figure 5.2). The ability of the SNX dimer to interact with either end of the Vps26/29/35 subcomplex has lead to the suggestion that two complexes may lie next to one another, forming a coat on a tubular vesicle structure. The structure of a relevant retromer associated sorting nexin is critical in building the entire picture of this macromolecular assembly. Further structures of individual retromer components, complexes of interacting partners and ultimately the entire retromer complex itself would be of great interest and use in furthering understanding of endosome-to-Golgi retrieval. Sufficient production of soluble SNX1 is essential in this process.

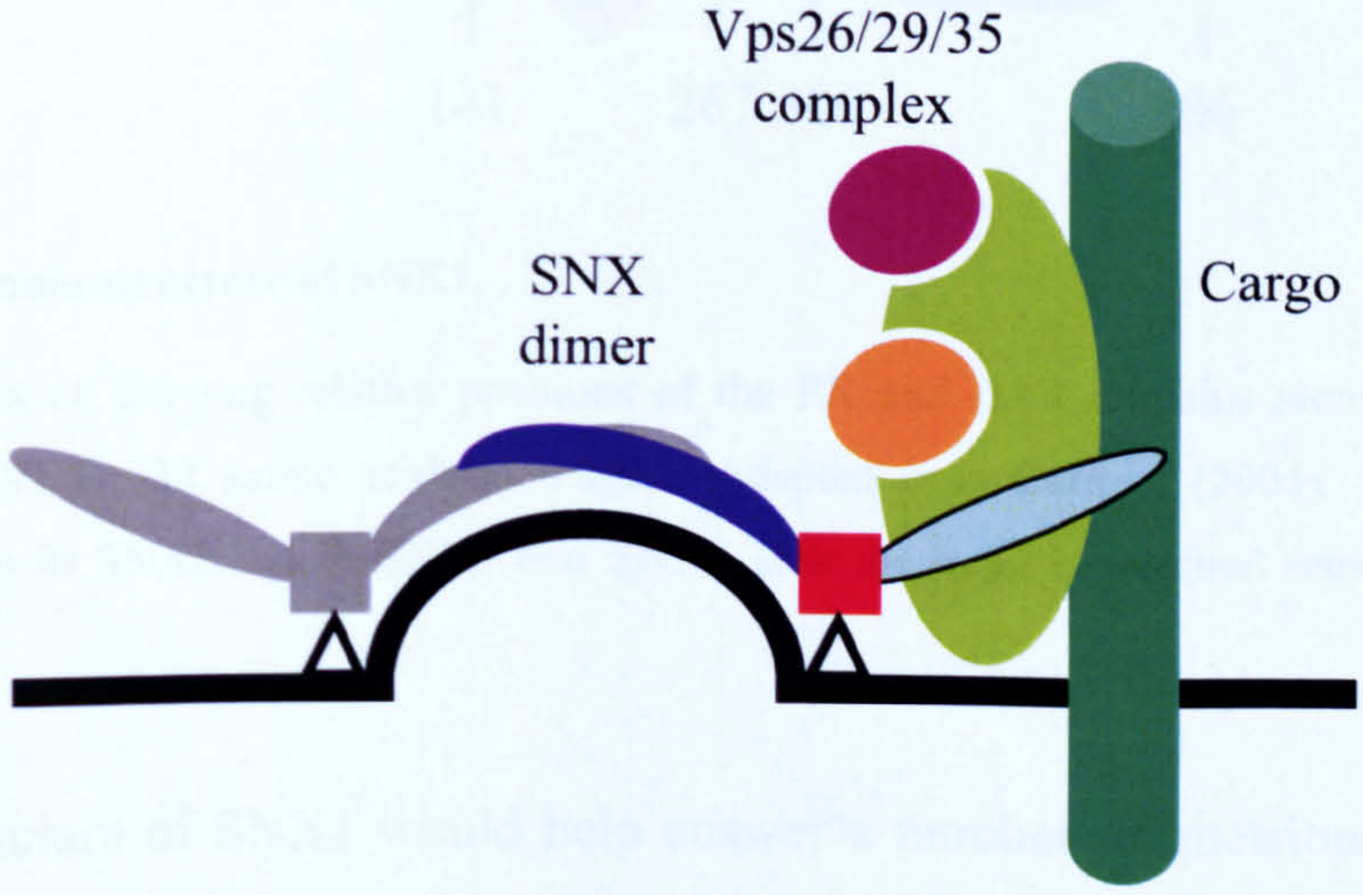


Figure 5.2 Schematic of the retromer complex.

A representation of the retromer components bound to membrane (black line). The coloured SNX monomer may be SNX1 or SNX2, while the grey monomer may consist of SNX1, SNX2, SNX5 or SNX6. PX domains (red) are shown interacting with phosphoinositide (triangles) and curved membrane is seen interacting with the Bar (blue) dimer interface. Cargo (dark green) interacts predominantly with Vps35 (light green). Vps29 (purple) binds at one end of Vps35, while Vps26 (orange) binds nearer the other. Adapted from Cullen, 2008.

5.1.4 Sorting nexin 1

Sorting nexin 1 is the most widely studied member of the sorting nexin family due to its role in the mammalian retromer (as described in 5.1.3). It does, however, also function in other trafficking processes independent of the retromer, although these are less well understood (Gullapalli *et al.*, 2006).

Both the BAR and PX domains (Figure 5.3) are required for the correct localisation of SNX1 to the early endosome, with the BAR domain providing specificity for highly curved membranes and the PX domain for PtdIns(3)P, both features of the early endosome (Carlton *et al.*, 2004). It is these features that also provide the targeting specificity of the retromer making understanding these interactions highly important.

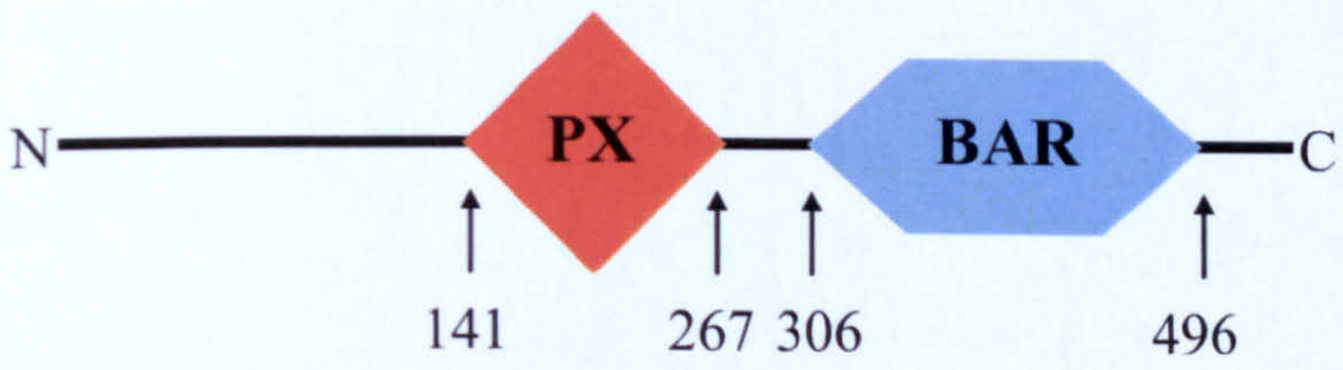


Figure 5.3 Domain structure of SNX1.

Schematic of SNX1 showing relative positions of the PX and BAR domains along with amino acid boundaries. SNX1 is 522 amino acids in length. Adapted from Carlton, (2004). A similar domain structure is seen in SNX2, while SNX5 and SNX6 lack the large N-terminal region (Carlton *et al.*, 2005a).

A crystal structure of SNX1 would help answer a number of questions, including the physical relationship between the PX and BAR domains, the nature of SNX1-PX novel dual specificity for both PtdIns(3)P and PtdIns(3,5)P_2 and why the SNX1-BAR domain appears to be highly sensitive to changes in curvature, whereas many other solved BAR domains are markedly less sensitive. Comparative studies with the SNX2 BAR domain would also be valuable, as only SNX1 induces membrane tubulation despite 69% sequence identity in this region (Carlton *et al.*, 2005b). A crystal structure of SNX1 would also form another important addition to the composite structure of components of the mammalian retromer. To initiate this structural work, the following experiments

where aimed at establishing soluble expression and purification procedures for SNX1 and important sub-regions within it. The production of crystallographic quality and quantity protein is the first step that will lead to the structural elucidation of this important protein.

5.2 Methods

The methods used in the study of SNX1 are presented below. Some methods from Chapter 2 were also used and will not be repeated here, but referenced by their section numbering where appropriate.

5.2.1 Full length SNX1

The pGEX-4T-3 vector containing the entire human SNX1 coding sequence (pGEX-SNX1) was supplied by Prof. P. Cullen (University of Bristol) as a glycerol stock of previously transformed *E. coli* BL21 (DE3). The pGEX-4T vector system adds a N-terminal GST tag followed by a thrombin cleavage sequence to the recombinant protein. The use of a GST tag is recognised to produce the soluble expression of proteins normally poorly soluble when poly-His tags are used (Braun *et al.*, 2002). The pGEX vector contains the ampicillin resistant gene, therefore 50μg.ml^{-1} of ampicillin was always present for selection pressure.

5.2.1.1 Expression

Expression was undertaken as previously optimised by Dr Gyles Cozier, University of Bristol. A single colony was used to inoculate 10ml LB (with ampicillin) and grown overnight at 37°C with shaking (300rmp). This was then used to inoculate 1L of LB and grown at 37°C to an OD_{600} of approximately 0.6. The culture was then transferred to a 16°C water bath to cool. Once chilled, the cultures were induced with 0.1mM IPTG and further grown at 16°C in a refrigerated incubator (Innova4230, New Brunswick Scientific). Expression occurred overnight (approximately 16hrs) and the resulting cells were pelleted by centrifugation at 10,000g, 4°C for 10 min, resuspended in 40ml PBS and re-centrifuged at 5000g, 4°C for 5 min. Cell pellets were frozen at –20°C until required.

5.2.1.2 GST column purification

The initial purification of GST tagged SNX1 was performed by resuspending a cell pellet in 30ml of GST Binding Buffer (50mM HEPES pH 8.0, 200mM NaCl, 1mM

EDTA, 1mM dithiothreitol (DTT), 10% glycerol). This also contained an EDTA-free protease inhibitor (Roche). Cell disruption was performed as described for GRHPR (section 2.3.11). Resulting cell debris and insoluble proteins were pelleted by centrifugation at 15,000g for 25 minutes.

The GST tag allows for protein purification via binding to the glutathione ligand. In this work, 1ml GSTrap FF columns (Amersham Biosciences) which contain glutathione immobilised on cross-linked 4% agarose were used. The GSTrap FF column was equilibrated in GST Binding Buffer and the clarified supernatant loaded onto the column with a peristaltic pump at 0.3ml.min^{-1}. The flow through was collected and the column washed to a flat A_{280} baseline. A gradient into 20mM reduced glutathione was run at 0.3ml.min^{-1} over 10 CV to elute bound protein, with 0.5ml or 1ml fractions collected. Fractions corresponding to A_{280nm} peaks were analysed by SDS-PAGE and those containing recombinant protein pooled.

The amount of protein was estimated using the predicted molecular weight of GST-SNX1 (85,459 Da) and theoretical extinction coefficient (90760 $M^{-1}cm^{-1}$) calculated from Protparm.

5.2.1.3 Removal of the GST tag

Cleavage trials to remove the GST tag with thrombin were performed. A small aliquot of purified protein was mixed with 10U of thrombin per mg of recombinant protein. This was left at 18°C and samples taken at various time points, SDS-PAGE sample loading buffer added and the sample boiled to halt the cleavage reaction. Cleaved samples were then analysed on SDS-PAGE and compared to undigested protein. Prior to larger scale digestions, protein was buffer exchanged into the appropriate buffer for the subsequent purification step.

5.2.1.4 Subsequent purification attempts

Two methods were undertaken to further purify full length SNX1. These were attempted both prior to, and after thrombin digestion to remove the uncleaved protein,

thrombin and free GST. Generally ion exchange was undertaken first and followed by gel filtration, however gel filtration was also attempted alone.

Ion Exchange

SNX1 was loaded onto an equilibrated MonoQ column (Amersham Biosciences) in binding buffer (50mM HEPES pH 7.0 or 8.0, 20mM NaCl, 1mM EDTA, 1mM β-mercaptoethanol) and elution proceeded with a gradient into 1M NaCl run at 1ml.min^{-1}. Eluted protein was collected in 1ml fractions and those corresponding to A_{280nm} peaks were kept and analysed by SDS-PAGE.

Gel Filtration

A Superdex 200 16/60 column (Amersham Biosciences) equilibrated in 50mM HEPES pH 7.0 or 8.0, 150mM NaCl, 1mM EDTA, 1mM β-mercaptoethanol was used for gel filtration purification. Fractions from ion exchange identified as containing SNX1 were concentrated to approximated 0.5ml and loaded onto the column. Alternatively, cleaved protein was concentrated and loaded directly. Gel filtration was then run at 0.3ml.min^{-1} with 2ml fractions collected. Those fractions corresponding to A_{280nm} peaks were kept and analysed by SDS-PAGE.

5.2.2 SNX1 truncation construction

5.2.2.1 Primer design

To reduce the size of the protein and potential crystallisation complications, various SNX1 truncations were designed and constructed. Firstly, the largely unstructured N-terminus of SNX1 was removed and the remaining PX-Bar fragment cloned into both GST (pGEX-4T-1, Amersham Biosciences) and poly-His (pET-30b, Novagen) tagged expression vectors. Additionally, a poly-His tag was added to the C-terminal of the GST-PX-Bar version in an effort to assist with purification. The individual PX and Bar domain were also cloned into the poly-His (pET-30b) tagged expression vector.

Primers were designed with similar features (Table 5.1), including beginning and ending in a G or C and having at least 21bp of gene specific sequence. Restriction enzyme sites were included to facilitate ligation into the selected vectors and artificial stop codons (UAA) incorporated where required. More specifically, the

SX1_C_His_reverse primer includes the sequence to create an artificial hexa-His tag at the SNX1 C-terminus. The forward primers for the separate domain constructs incorporate a Factor Xa cleavage sequence (IEGR) directly at the protein N-terminus and the reverse primers contain an artificial stop codon. These are required as the pET-30b vector produces a N-terminal poly-His tag which, after thrombin cleavage, results in a 21 amino acid N-terminal extension. Also, this vector would normally include a non-cleavable C-terminal His-tag. The sequence of the primers used and their individual attributes are given in Table 5.1. All primers were produced and purified by Invitrogen and the lyophilised primers resuspended in Elgastat water to a final concentration of 0.1mM and stored at -20°C until required.

pGEX-4T-1
PXSX1_forward (b, c)
5′CGGAATTCCAGGAGGATCAATTTGATTTGACAG3′
Full_SX1_rev (b)
5′GCGCGGCCGCTTAGGAGATGGCCTTTGCCTCAGG3′
SX1_C_His_reverse (c)
5′GCGCGGCCGC**TTA**GTGGTGGTGGTGGTGGTGGGAGATGGCCTTTGCCTCAGG3′
pET-30b
PX_forward (d, e)
5′GCAGATCTGATCGAAGGTCGTCAGGAGGATCAATTTGATTTGACAG3′
PX_reverse (e)
5′GCGCGGCCGC**TTA**CTCTTCTTTTTCCAAGAACTCTCTGAC3′
BAR_forward (f)
5′GCAGATCTGATCGAAGGTCGTAGCAAAATGACCATCAAGATGAATG3′
BAR_reverse (d, f)
5′GCGCGGCCGC**TTA**TGAGTAAAGGAGTGTCTCAAGGTAC3′

Table 5.1 Primers designed for the cloning of SNX1 truncations.

The pGEX-4T-1 vector was used for the GST tagged truncations, while pET-30b was used for N-terminal poly-His tagged truncations. All primers are given in the 5' to 3' direction. Incorporated restriction endonuclease sites are indicated; orange: *EcoR*I, red: *Not*I, green: *Bgl*II. Sequences for the addition of Factor Xa protease sites (blue), hexa-His tag (purple) and stop codons (bold) are shown. Brackets indicate which truncations they were used to create (Figure 5.8).

5.2.2.2 PCR and vector construction

The pGEX-SNX1 plasmid DNA was extracted from 5ml of a 10ml overnight culture of the supplied *E. coli* BL21 (DE3) strain using the QIAprep Spin Miniprep Kit (Qiagen), following manufacturers' instructions. This was used as the template plasmid for the PCR of the GST tagged PX-Bar construct (GST-PX-Bar); subsequent PCR used this truncated construct as the template. The standard PCR reaction mix was used in all cases: 1.75U Expand High Fidelity Polymerase, 1x Expand Buffer + $MgCl_2$, both primers (as indicated in Table 5.1) at 2μM, dNTPs (50pmol) and 1μl of template plasmid, in a final volume of 50μl. The PCR cycling protocol varied depending on the primers used. The basic protocol, as used in the PCR of the GST-PX-BAR construct, is given in Table 5.2 with variations described below.

	Temperature	Time	No. Rounds
Initial DNA denaturation	94°C	4 minutes	1
Denaturation	94°C	1 minute	25
Annealing	48°C	45 seconds	
Amplification	72°C	90 seconds	
Hold	4°C	Indefinite	

Table 5.2 General PCR protocol for SNX1 truncations.

Due to the GC rich His-tag sequence in the *SX1_C_His_reverse* primer it had a much higher melting point (Tm = 75°C) then its complement primer (*PX_SX1_forward*, Tm = 57°C). The difference in Tm of 18°C for this primer pair required a two-step annealing process for PCR of the GST-PX-Bar-His construct to be successful. During the first 5 rounds of amplification, primer annealing occurred at 48°C thus allowing a certain amount of non-specific hybridisation of the reverse primer. During the subsequent 20 rounds the annealing temperature was increased to 55°C, with the previous amplification in target copy number increasing the chance of the forward primer successful binding and raising the stringency of the reverse primer. The protocol for the poly-His tagged PX-Bar construct (His-PX-Bar) required primer annealing to occur at 55°C across all rounds. The PCR cycle for the His-tagged PX (His-PX) and Bar (His-Bar) individual domains was as in Table 5.2 except with an amplification time of 60 seconds as befits their smaller size.

The subsequent purification, cloning, ligation, transformation and sequencing was performed largely as described in sections 2.3.3 to 2.3.8 with the substitution of appropriate restriction enzymes and expression vectors as indicated in Table 5.1. The creation of the expression constructs was done by double digestion of both the expression vector and corresponding pGEM-Teasy ligation. The vector pGEX-4T-1 and the inserts designed for this vector (GST-PX-Bar and GST-PX-Bar-His) were cleaved with *EcoR*I and *Not*I, while pET-30b and the inserts intended for this (His-PX-Bar, His-PX and His-Bar) were digested with *Bgl*II and *Not*I. Overnight ligation was followed by transformation into *E. coli* NovaBlue, Rosetta and BL21 (DE3) strains with final double digests to confirm the presence of the insert. Plasmid retention required ampicillin (50μg.ml^{-1}) for pGEX and kanamycin (50μg.ml^{-1}) for pET vectors. Correct colonies were then grown for expression trials and long term -80°C storage.

5.2.3 GST-PX-Bar

It was hoped that removal of the unstructured N-terminal section would improve the purification and stability of the resulting protein. Expression trials were run as described in section 2.3.10. Expression temperatures of 16°C and 37°C were investigated, with IPTG concentrations of 1mM and 0.1mM. Final large scale expression and purification was undertaken as previously described for full length SNX1 (section 5.2.1). Thrombin cleavage and further purification was also as above, except purification buffers contained 10% v/v glycerol and 0.1mM phenylmethylsulphonyl fluoride (PMSF).

5.2.4 Poly-His tagged PX-Bar constructs

Expression trials were performed as described in section 2.3.10. Expression temperatures of 16°C, 30°C and 37°C and IPTG induction concentrations of 1mM and 0.1mM were trialled.

5.2.5 Denaturing nickel purification

Expression of both truncation constructs GST-PX-Bar-His and His-PX-Bar resulted in insoluble protein. Cell disruption was done as described for GRHPR (section 2.3.11), with the cell pellet resuspended in 30ml of 50mM HEPES pH 8.0, 500mM NaCl, 20mM imidazole, 10% (v/v) glycerol, 1mM β-mercaptoethanol and an EDTA-free

protease inhibitor (Roche). After centrifugation at 25000g for 30min the recombinant protein was present in inclusion bodies. The insoluble pellet was resuspended in Denaturing SNX Ni^{2+} Loading Buffer (50mM HEPES pH 8.0, 500mM NaCl, 20mM imidazole, 8M urea, 10% (v/v) glycerol and 1mM β-mercaptoethanol). A sonicating water bath was used to assist the resuspension of denatured protein. The denatured protein was clarified by centrifugation as above. The solubilised, denatured protein was loaded on to a 5ml Ni^{2+} affinity column equilibrated in the same buffer and washed until a flat A_{280nm} reading was achieved. Bound protein was then eluted using a gradient to 1M imidazole. The fractions containing denatured protein were identified

5.2.6 Refolding GST-PX-Bar-His

In both cases the success of refolding was analysed by solubility and successful thrombin cleavage.

5.2.6.1 Dialysis buttons

Against a background of 250mM NaCl, 10% glycerol and 1mM β-mercaptoethanol, a range of buffer pHs were trialled to assist refolding of this construct (Table 5.3). A small volume (30μl) of purified, denatured GST-PX-Bar-His was added to a dialysis button and sealed with a square of dialysis tubing (14,000 Da cutoff). Dialysis was undertaken with 50ml of "refolding buffer" at 4°C. After 16 hrs the sample was inspected for precipitation, removed from the button and centrifuged at 18,000g for 10min. The resulting supernatant was analysed by SDS-PAGE for the presence of protein.

Buffer (50mM)	pH
Sodium Acetate	4.6
Sodium Citrate	5.6
BisTris-HCl	6.5
Potassium phosphate	7.5
Tris-HCl	8.5
Ches	9.5

Table 5.3 Various buffers attempted during GST-PX-Bar-His refolding.

5.2.6.2 Refolding screen

Small scale trials using a commercially available protein refolding screen (MDL) were undertaken. This screen contains 15 pre-made buffers containing various combinations of the buffer components shown in Table 5.4. Denatured protein (1mg.ml^{-1}) was diluted 1:20 into these buffers with rapid mixing. These were then incubated for 1hr at 18°C, with continued slow mixing. Protein is initially determined to be soluble (and by inference refolded) by its visible presence on an SDS-PAGE gel after centrifugation at 18000g, 15°C for 30min. Refolding success could then be scored as a comparison to the original denatured protein and critical buffer components determined.

Buffer	50mM MES pH 6.0 50mM Tris-HCl pH 8.0
Salt	9.6mM NaCl, 0.4mM KCl 240mM NaCl, 10mM KCl 2mM $MgCl_2$, 2mM $CaCl_2$
Redox Potential	+/- 1mM DTT +/- 1mM GSH, 0.1mM GSSH
Detergent	0.5% Triton X-100 0.05% polyethylene glycol 3,550
Solubiliser	0.4M sucrose 0.5M arginine 0.75M guanidine HCl

Table 5.4 Components of the MDL refolding screen.

Each of the components was present in a number of the pre-made buffers allowing their contribution to refolding to be determined.

5.2.7 Poly-His tagged domain constructs

5.2.7.1 Expression and purification

Expression trials were run as described in section 2.3.10 with expression temperatures of 30°C and 37°C and IPTG concentrations of 1mM and 0.1mM. Large scale expression was undertaken at 37°C with 1mM IPTG for 3 – 5 hrs and cells pellets

prepared using standard procedures. Initial nickel column purifications were also performed as for T-GRHPR (section 2.3.11).

5.2.7.2 His-tag removal

Prior to protease cleavage, purified protein was buffer exchanged into the appropriate buffer for the subsequent purification step. Cleavage trials were performed with both thrombin and Factor Xa to determine the most appropriate cleavage site, digestion time and temperature. Both proteases were added initially at 10U or 5U per mg of protein. After cleavage, the serine protease and uncleaved protein were removed by benzamidine and Ni^{2+} column purification as described in section 2.3.13.

5.2.7.3 Gel filtration

Gel filtration was done largely as described in section 2.3.14. Both the PX and Bar domains were purified via a HiLoad 16/60 Superdex 75 prep grade column. The PX domain was purified in 20mM Ches pH 9.5 and 200mM NaCl buffer and the Bar domain in 20mM Ches pH 9.5, 200mM NaCl, 10% glycerol and 1mM β-mercaptoethanol buffer. Fractions (1ml) corresponding to A_{280} peaks were analysed by SDS-PAGE.

5.2.7.4 Crystallisation

Purified domains were buffer exchanged into either 20mM Tris-HCl pH 8.5 (PX) or 20mM Tris-HCl pH 8.5 and 1mM β-mercaptoethanol (Bar) and concentrated as described previously. A range of final protein concentrations from 2.4 – 16.7mg.ml^{-1} for the PX domain and 1.5 – 10.2mg.ml^{-1} for the Bar domain were used in crystallisation trials with a number of commercially available screens (Table 5.5). Crystallisation screens were set up as sitting drops with the Phoenix nanolitre robot using 250nl drops of mother liquor and 250nl drops of protein. Plates were sealed, stored at 18°C and inspected for indications of crystal growth at regular intervals. In addition, JBScreens 3, 5, 6 and 7 were set up manually as 500nl mother liquor + 500nl protein hanging drops in pre-greased 24-well plates as described in section 2.3.15.2.

Phosphatidylinositol complex attempt

The PX domain of SNX1 is thought to bind 2 different phosphatidylinositides (PtdIns), PtdIns(3)P and PtdIns(3,5)P_2 (Cozier *et al.*, 2002). Solid di-butanoyl-phosphatidylinositidol 3-phosphate (di-C4-PtdIns(3)P) and di-butanoyl-phosphatidylinositidol 3,5-bisphosphate (di-C4-PtdIns(3,5)P_2) were supplied by Echelon, Utah. The production of these had been discontinued and only a small amount could be purchased limiting the screening that could be done. The powder was resuspended in water to a concentration of 30mM and stored at -20°C. An aliquot of this resuspension (3.3µl) was added to concentrated PX domain (46.7µl, 5.1mg.ml^{-1}) to give a final concentration of 2mM PtdInsP. This was incubated on ice for 30 minutes, followed by centrifugation at 18,000g for 5 minutes. The Phoenix robot was then used to set up 250nl of Structure Screen I and II HT-96 with 250nl of "complex". Plates were sealed, stored at 18°C and inspected for crystal growth as usual.

Crystal Screen	**Manufacturer**	**No. of conditions**
Structure Screen I and II HT-96	MDL	96
PACT *premier* HT-96	MDL	96
JCSG-*plus* HT-96	MDL	96
JBScreen 1 – 4	Jena Bioscience	96
JBScreen 5 – 8	Jena Bioscience	96
JBScreen 9 and 10	Jena Bioscience	48

Table 5.5 Commercial crystallisation screens used when screening SNX1 truncations.

5.3 Results

5.3.1 Overview

In this chapter work undertaken to purify full length human SNX1 will be presented. In addition the cloning and expression of a number of SNX1 truncation constructs, produced in the hope of improving purification and crystallisability is reported. The crystallisation trials undertaken on the individual SNX1 PX and Bar domains, including co-crystallisation with phosphatidylinositides, are also presented. Disappointingly, purification complications and a lack of reproducible crystallisation hits prevented further progression of this work. However, the establishment of expressing constructs and purification procedures will facilitate future work.

5.3.2 Full length SNX1

5.3.2.1 Purification

The established procedure of expression at 16°C induced with low (0.1mM) IPTG concentration resulted in the soluble expression of full length GST-tagged SNX1. Initial purification of this protein was via the interaction between the GST tag and immobilised glutathione as described in 5.2.1.2.

Protein bound to the column was washed with a gradient from 0 - 20mM glutathione to competitively disrupt the GST:resin interaction, with protein elution monitored by A_{280} measurements. The only peak eluted at approximately 4mM glutathione corresponding to reasonably pure protein as determined by SDS-PAGE (Figure 5.4 and Figure 5.5). Although the size of the eluted protein cannot be judged accurately due to it exceeding the limit of the molecular weight markers used, the dominant band is likely to correspond to the 85kD GST-SNX1 monomer (Figure 5.5). The fractions containing the recombinant protein were pooled, yielding less than 2mg of protein from a litre culture, as calculated with the extinction coefficients given in Table 5.6.

A significant amount of the expressed GST-SNX1 did not bind the GST column and can be seen in the flow-through fraction produced during column loading (Figure 5.5). Attempts were made to increase the proportion of GST-SNX1 to bind the column.

Temperature, flow rate and buffer pH are all thought to affect the GST-tag/glutathione resin interaction (GSTrap FF Troubleshooting, Amersham Biosciences). To investigate this, the loading of the GST column was repeated at pH 7.0, the suggested optimal loading pH (GSTrap FF Troubleshooting, Amersham Biosciences). Additional binding trials were also done at a slow flow rate (0.1ml.min^{-1}) and 4°C, both conditions suggested to improve binding affinity. No improvement in binding yield was seen with any altered protocol.

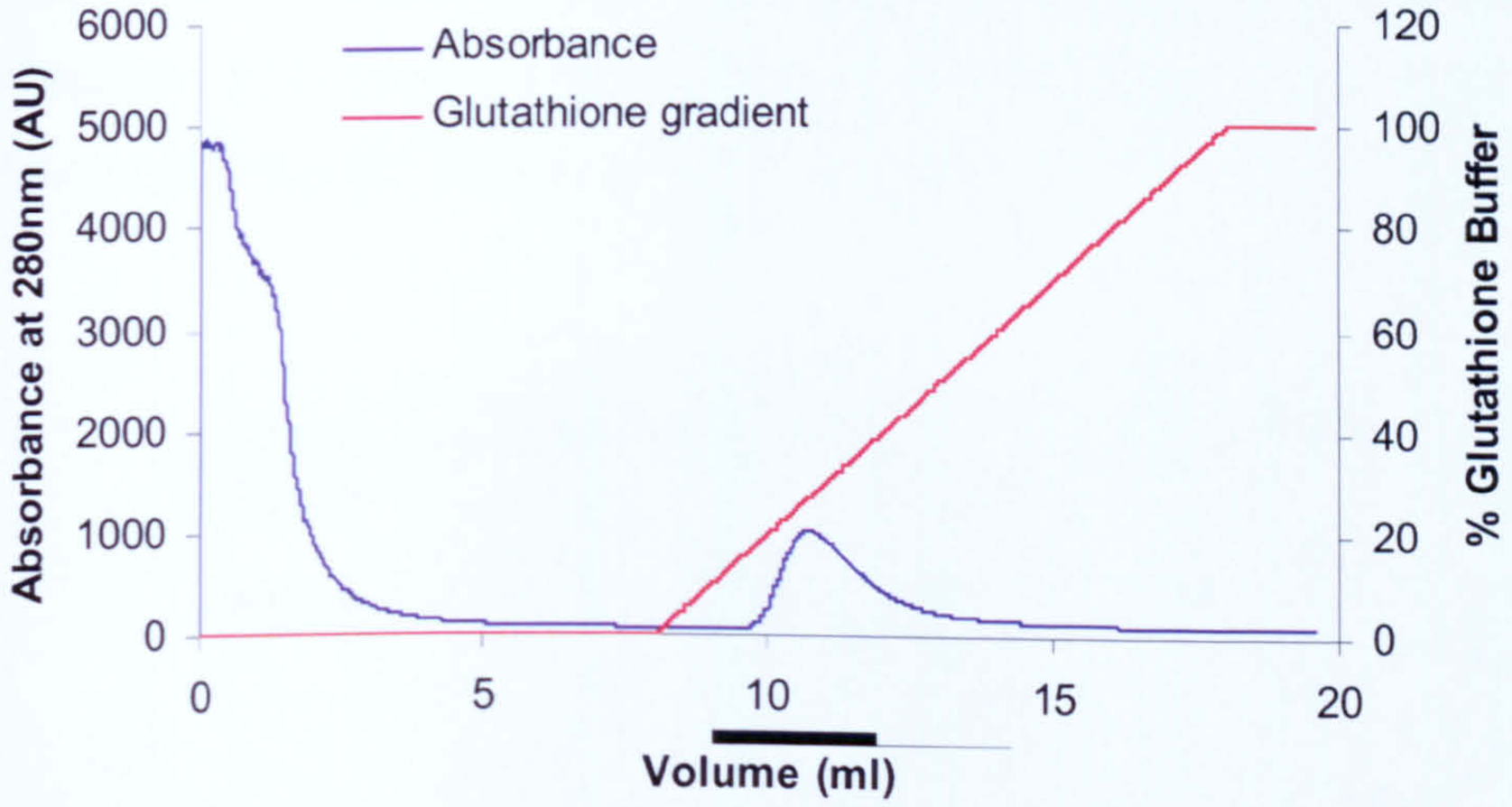


Figure 5.4 Glutathione affinity column chromatography of full-length SNX1.

Profile of GST-SNX1 elution with a glutathione gradient, monitored by UV absorbance at 280nm. The solid line indicates those fractions (0.5ml) ran on the SDS-PAGE gel in Figure 5.5.

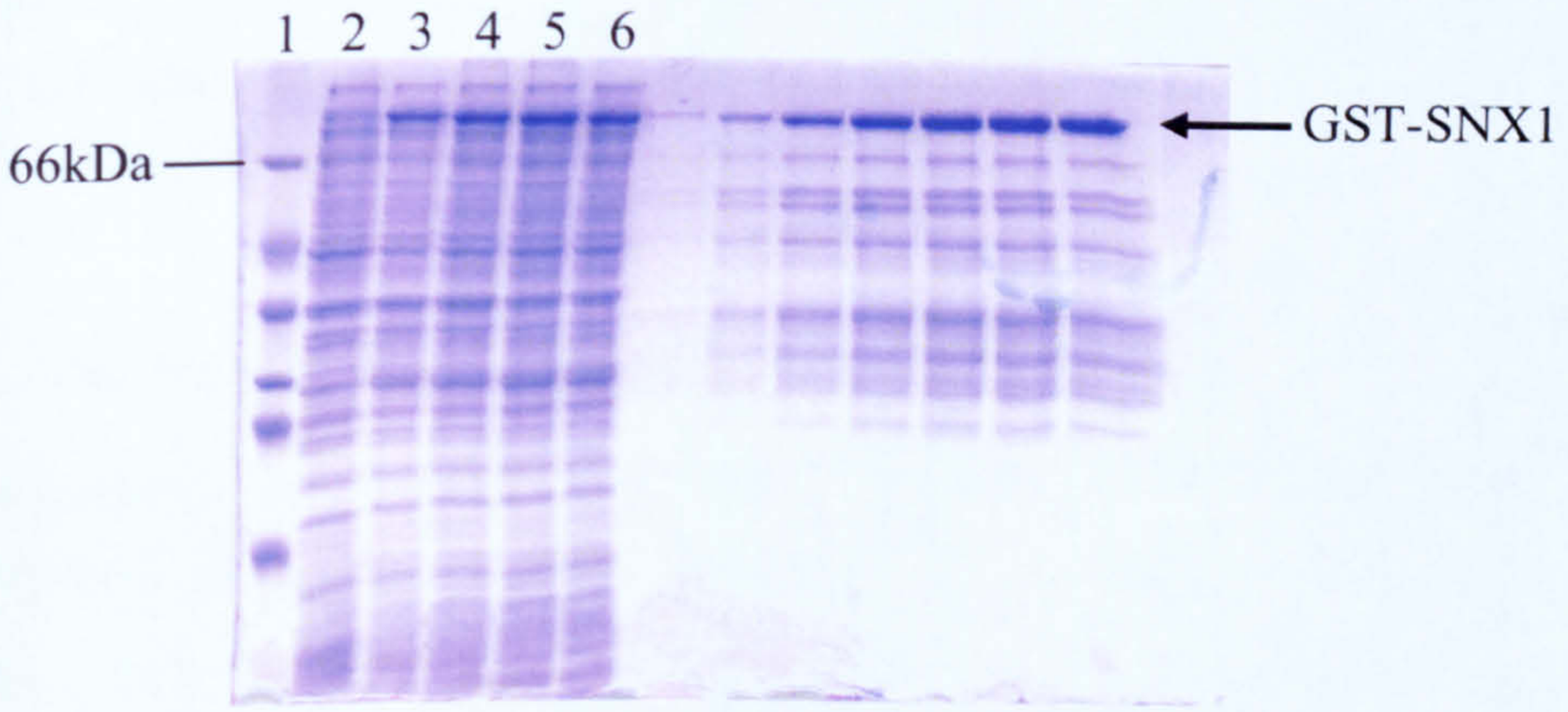


Figure 5.5 SDS-PAGE of purification fractions from GST-SNX1.

Purification of full length SNX1; lane 1) molecular weight markers, 2) uninduced total protein, 3) total protein after 8hr induction, 4) total protein after overnight induction, 5) soluble protein after overnight induction, 6) flow through fraction from glutathione column chromatography. Other lanes correspond to fractions marked with a solid line in Figure 5.4.

5.3.2.2 GST tag cleavage

The pooled protein from the GST column step was subjected to thrombin digestion (10U/mg) at 18°C in order to remove the GST tag. A time course showed complete cleavage had occurred after 19hrs, resulting in an upper SNX1 band, running slightly higher then its expected 59kDa, and a lower 26kDa GST band (Figure 5.6). Removal of the GST tag by thrombin cleavage results in 5 non-native amino acids (GSPNS) remaining at the SNX1 N-terminus. Protein sequencing on the upper band, by the University of Bristol Proteomics Facility, produced the sequence GSPNSMAS. This corresponds to the expected 5 amino acid tag remnant plus the first 3 residues of native SNX1, confirming the identity of the protein and correct N-terminal cleavage.

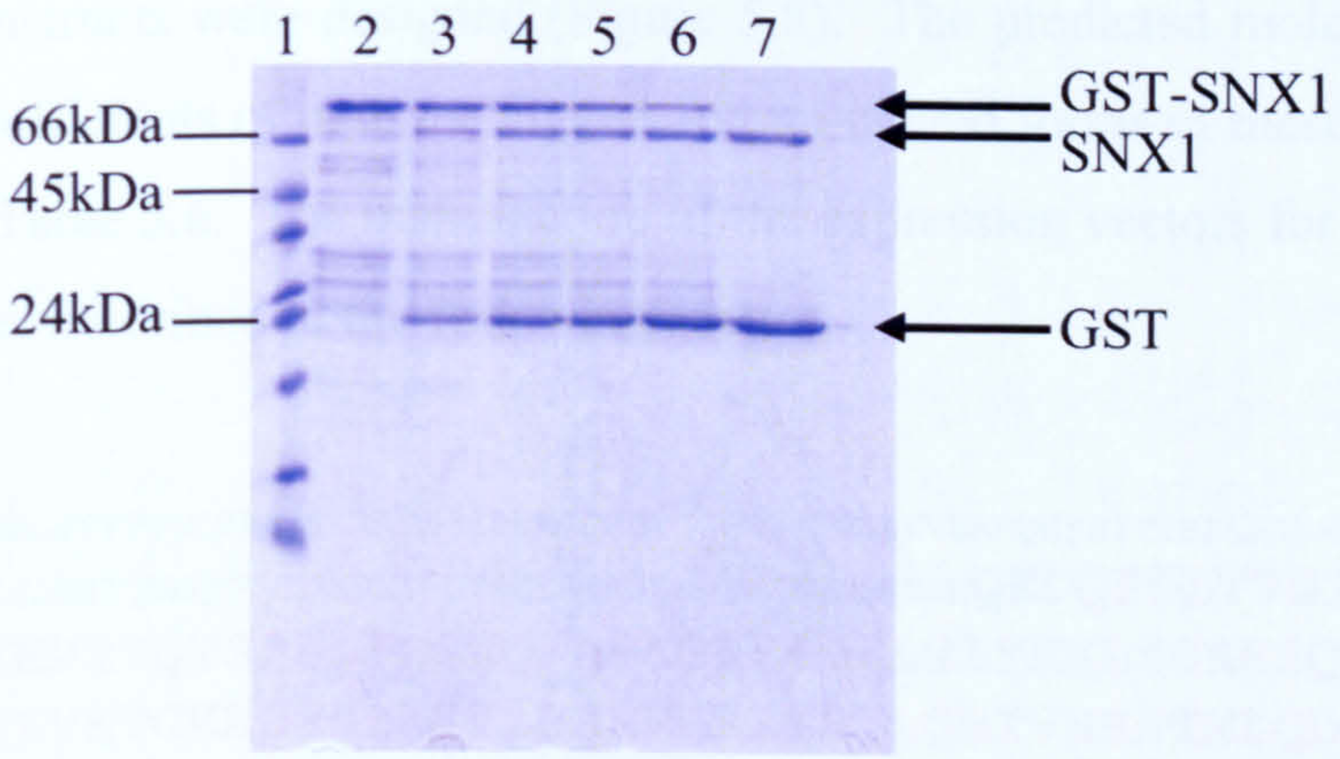


Figure 5.6 SDS-PAGE gel of thrombin cleavage trial undertaken on GST-SNX1.

Thrombin cleavage used to remove the GST tag from the full length SNX1 protein purified in Figure 5.5; lane 1) molecular weight markers, 2) GST-SNX1 pre-thrombin addition, 3) 1hr post thrombin addition, 4) 2hrs, 5) 3hrs, 6) 8hrs, 7) 19hrs. Major cleavage products are labelled.

5.3.2.3 Further purification attempts

Any attempts to further purify the cleaved full length SNX1 by either ion exchange or gel filtration as described in the methods (5.2.1.4) resulted in complete loss of the protein. This suggested that the removal of the GST tag decreased the stability of SNX1. Attempts to further purify the uncleaved, GST-tagged form by the same methods were also unsuccessful. The protein behaved as large impure aggregates in gel filtration chromatography and extensive protein loss was observed in ion exchange chromatography, which prevented further progress with this construct (data not shown).

5.3.3 SNX1 truncation constructs

While a small amount of the full length GST-SNX1 protein could be produced and initially purified, the loss of protein during subsequent purification steps made it impossible to carry out further work upon this construct. The domain structure of the full length SNX1 protein had been defined previously (Carlton, 2004) and is reproduced in Figure 5.7. This was analysed to determine potential protein truncations that could be constructed to improve the chances of crystallisation while still answering some of the questions regarding SNX1. Using previously published work including the construct design of the SNX1-PX domain by Zhong *et al.*, (2005) used for NMR studies and the related dAmphiphysin (dAmp) Bar structure from Peter *et al.*, (2004) along with helpful discussions with Dr R. Sessions (University of Bristol), a number of truncated constructs were designed (Figure 5.8). The predicted molecular weight and extinction coefficients of both the tagged and un-tagged forms of these novel constructs are given in Table 5.6. The manufacture of the expression vectors for these truncations will be further described in the following section.

MASGGGGCSASERLPPPFPGLEPESEGAAGGSEPEAGDSDTEGEDIFTGAAVVSKHQSPKITTSLLPINNGSKENGIHEEQDQE
PQDLFADATVELSLDSTQNNQKKVLAKTLISLPPQEATNSSKPQPTYEELEEEE**QE**DQFDLTVGITDPEKIGD
GMNAYVAYKVTTQTSLPLFRSKQFAVKRRFSDFLGLYEKLSEKHSQNGFIVPPPP
EKSLIGMTKVKVGKEDSSSAEFLEKRRAALERYLQRIVNHPTMLQDPDVREFLEK
EELPRAVGTQTLSGAGLLKMFNKATDAV**SKMTIKMNES**DIWFEEKLQEVECEEQR
LRKLHAVVETLVNHRKELALNTAQFAKSLAMLGSSEDNTALSRALSQLAEVEEKI
EQLHQEQANNDFFLLAELLSDYIRLLAIVRAAFDQRMKTWQRWQDAQATLQKKRE
AEARLLWANKPDKLQQAKDEILEWESRVTQYERDFERISTVVRKEVIRFEKEKSK
DFKNHVIKY**LETLLYS**QQQLAKYWEAFLPEAKAIS

Figure 5.7 SNX1 protein sequence and domain boundaries.

Small font size depicts the putative N-terminal "unstructured" region. Red delineates PX domain, with bold extension included in PX constructs based on Zhong *et al.*, (2005). Blue demarcates the Bar domain as determined in Carlton (2004), with bold extensions included in Bar constructs due to comparison with the ordered region of the dAmp Bar structure (Peter *et al.*, 2004) and discussion with Dr R Sessions (University of Bristol). This also corresponds to the schematic of SNX1 in the introduction (Figure 5.3).

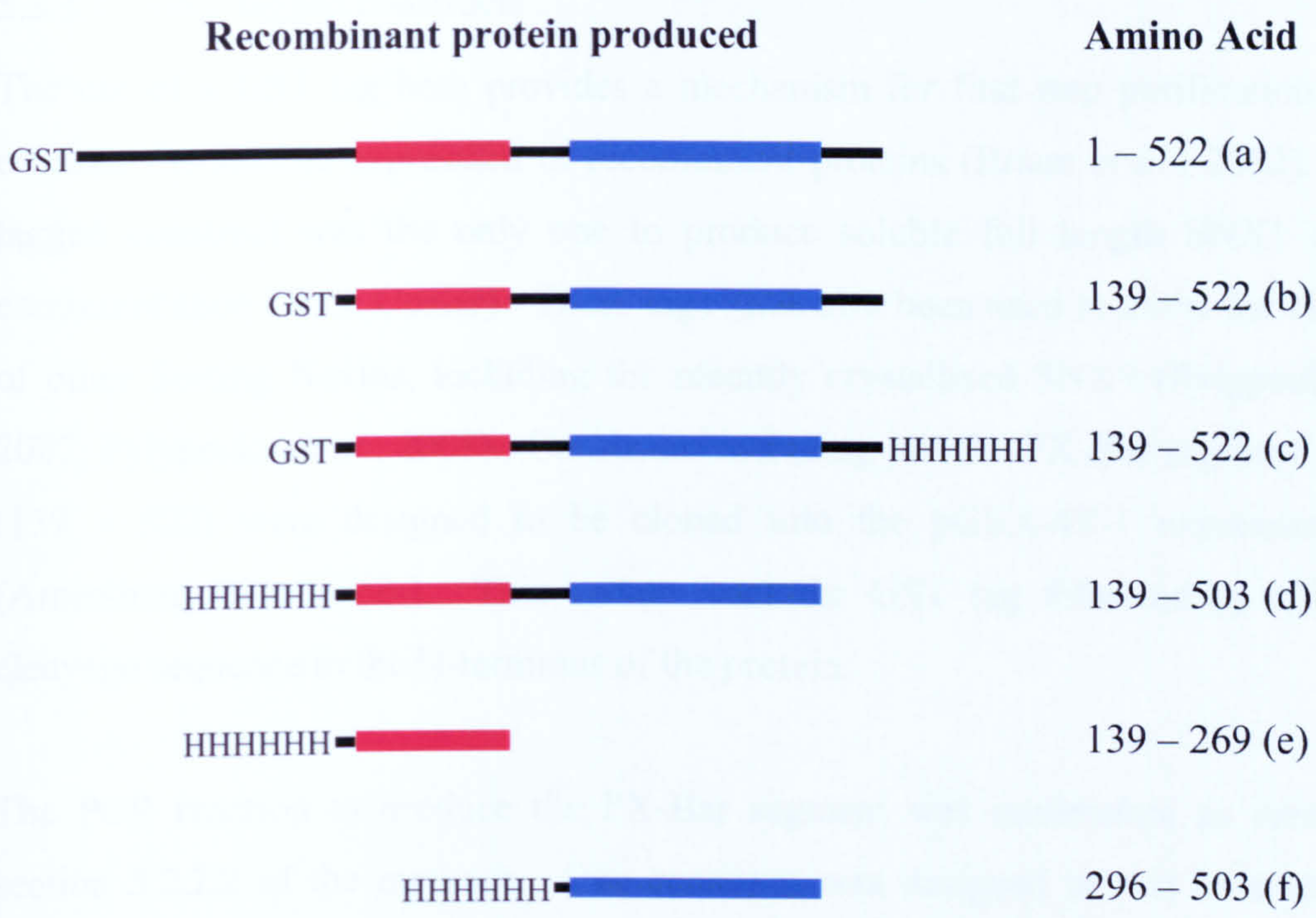


Figure 5.8 Schematic of SNX1 truncations including tags.

The positions of affinity purification tags are indicated; GST: glutathione S-transferase tag, HHHHHH: poly-His tag. Colored regions correspond to those in Figure 5.7 (and Figure 5.3) and amino acid numbers refers to the native SNX1 numbering. Letter codes; a) GST-SNX1, b) GST-PX-Bar, c) GST-PX-Bar-His, d) His-PX-Bar, e) His-PX, f) His-Bar. These letter also correspond to the primers used in their creation (Table 5.1)

Protein	MW (Da)	ε ($M^{-1}cm^{-1}$)
GST-SNX1	85459.1	90760
SNX1	59310.7	47900
GST-PX-Bar	71261.8	89520
PX-Bar	45113.4	46410
His-PX	19514.7	5960
PX	17896.3	5960
His-Bar	28982.9	33460
Bar	24716.1	33460

Table 5.6 Predicted molecular weights and extinction coefficients of designed SNX1 truncations.

Values used in all calculation of protein concentration. Calculated with Protparam (Expasy; http://ca.expasy.org).

5.3.3.1 GST tagged constructs

The use of a GST tag both provides a mechanism for first step purification and can enhance the soluble expression of recombinant proteins (Braun *et al.*, 2002). A GST tagged construct was the only one to produce soluble full length SNX1 (personal communication Dr G. Cozier). These tags have also been used to assist the expression of other Sorting Nexins, including the recently crystallised SNX9 (Pylypenko *et al.*, 2007; Pylypenko *et al.*, 2008). Constructs including just the PX-Bar segment of SNX1 (139 – 522) were designed to be cloned into the pGEX-4T-1 expression vector (Amersham Biosciences). This vector fuses the GST tag followed by a thrombin cleavage sequence to the N-terminus of the protein.

The PCR reaction to produce the PX-Bar segment was undertaken as described in section 5.2.2.2 of the methods. One construct was designed to just have the pGEX vector encoded GST tag, therefore PCR was performed using the primers *PXSX1_forward* and *Full_SX1_rev* (Table 5.3, Figure 5.9a). It was found that an extended time of 90 seconds was required for successful PCR and this time was used for all subsequent PX-Bar reactions. A second construct was designed to include a non-cleavable, C-terminal poly-His tag and therefore used primers *PXSX1_forward* and *SX1_C_His_reverse* (Table 5.3 and Figure 5.9b). This His-tag was included with the hope of improving purification by allowing the use of the tighter affinity nickel column purification, without the loss of the GST solubilising effect. Successful PCR for this construct required the modification of the annealing step as the two primers had significantly different Tm points (75°C and 57°C respectively). The use of lower stringency in the first 5 annealing rounds as described in the methods (5.2.2.2) was essential for the production of any PCR product, but is likely to be the cause of the smearing of DNA product (Figure 5.9b).

In both cases a PCR product of approximately 1.2 kbp was visualised in a 1% agarose gel (Figure 5.9a, b) and purified. Care was taken to avoid as much of the smeared region as possible present in Figure 5.9b when excising that product. Following ligation in to pGEM-Teasy and transformation into *E. coli* NovaBlue, a significant proportion of colonies were the expected white of successfully ligated plasmid:PCR product. A number of white colonies for each insert were selected and plasmid DNA

was purified. Double digestion using *EcoR*I and *Not*I confirmed the presence of the PCR product insert, and DNA sequencing verified that no mutations had been created. Full digestion of the pGEM-PCR vector with the *EcoR*I and *Not*I restriction enzymes produced a cleaved product which was then ligated into similarly digested pGEX-4T-1 as described in 5.2.2. Transformation into the *E.coli* Rosetta strain and selection on chloramphenicol and ampicillin LB agar plates resulted in a small number of colonies. Conformation of successful ligation was seen with a double digestion of purified plasmid (Figure 5.9c). These constructs are referred to as GST-PX-Bar and GST-PX-Bar-His and correspond to SNX1 truncations b and c in Figure 5.8.

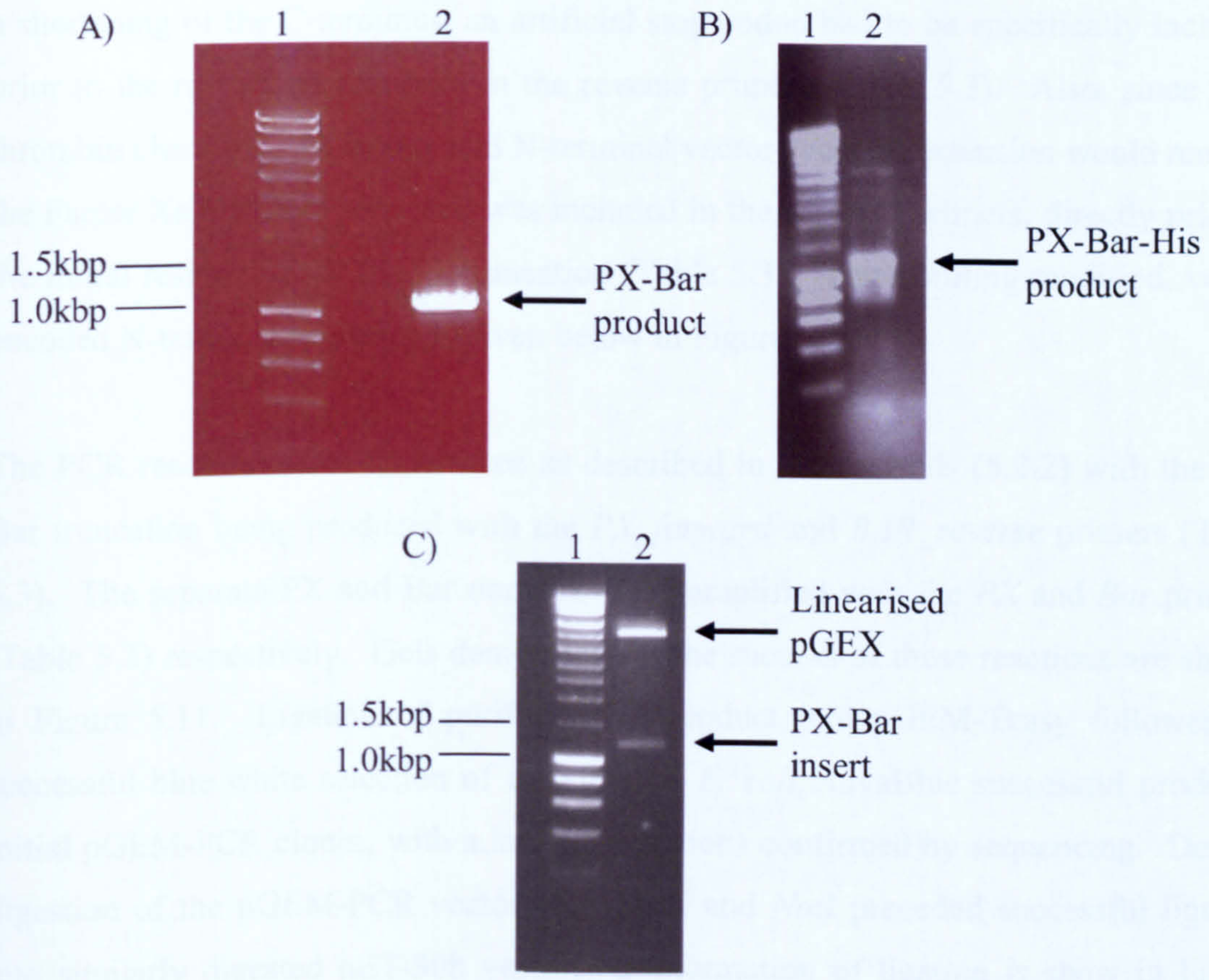


Figure 5.9 Agarose gels of GST-tagged PX-Bar vector construction.

1% agarose gels of; Lane 1) DNA ladder with relevant sizes indicated, A) PCR product PX-Bar (2), B) PCR product PX-Bar-His (2), C) example double vector digestion showing presence of insert (2). All show the correct sized insert of approximately 1.2kbp. Linearised pGEX vector is approximately 5kbp. All inserts were also confirmed by DNA sequencing.

5.3.3.2 Poly-His tagged constructs

Nickel affinity purification with the use of a poly-His tag allows for both faster column loading and more stringent on-column washing then GST tag purification, due to the increase affinity of the interaction (Schmitt, Hess and Stunnenberg, 1993). As the requirement of the SNX1 truncations for the solublising effect of the GST was unknown, a further 3 constructs were designed to be cloned into the pET-30b expression vector (Novagen). This vector adds a hexa-His tag and thrombin cleavage sequence to the N-terminus of the protein via a 18 amino acid linker. Due to the particularities of this vector and the truncations to be inserted into it, a number of specific additions were made to the primers. Firstly, as these truncations also involved a shortening of the C-terminus, an artificial stop codon had to be specifically included prior to the restriction sequence in the reverse primers (Table 5.3). Also, since after thrombin cleavage a 20 amino acid N-terminal vector encoded extension would remain, the Factor Xa cleavage sequence was included in the forward primers, directly prior to the initial residue of the SNX1 truncation (Table 5.3). The resulting modified, vector encoded N-terminal sequence is given below in Figure 5.10.

The PCR reactions were undertaken as described in the methods (5.2.2) with the PX-Bar truncation being produced with the *PX_forward* and *BAR_reverse* primers (Table 5.3). The separate PX and Bar domains were amplified with the *PX* and *Bar* primers (Table 5.3) respectively. Gels demonstrating the success of these reactions are shown in Figure 5.11. Ligation of purified PCR product into pGEM-Teasy followed by successful blue-white selection of transformed *E. coli* NovaBlue successful produced initial pGEM-PCR clones, with a lack of mutations confirmed by sequencing. Double digestion of the pGEM-PCR vector with *Bgl*II and *Not*I preceded successful ligation into similarly digested pET-30b vector. Conformation of ligation is show in Figure 5.11, with these constructs being referred to as His-PX-Bar, His-PX and His-Bar respectively (Figure 5.8 d, e, f). All constructs were finally transformed into the *E. coli* Rosetta strain for recombinant protein expression.

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPIEGR-**SNX1 truncation**

Figure 5.10 N-terminal protein sequence produced by the modified pET-30b vector tag.

The modified tag sequence included in the forward primers (Table 5.1) encoded a Factor Xa cleavage site (IEGR; blue), used for all N-terminally His-tagged SNX1 constructs shown in Figure 5.8 (d, e, f). Cleavage with this protease occurs directly after 'R' in this sequence. Thrombin cleavage occurs between the 'R' and 'G' of its recognition site (red) which is encoded by the original vector.

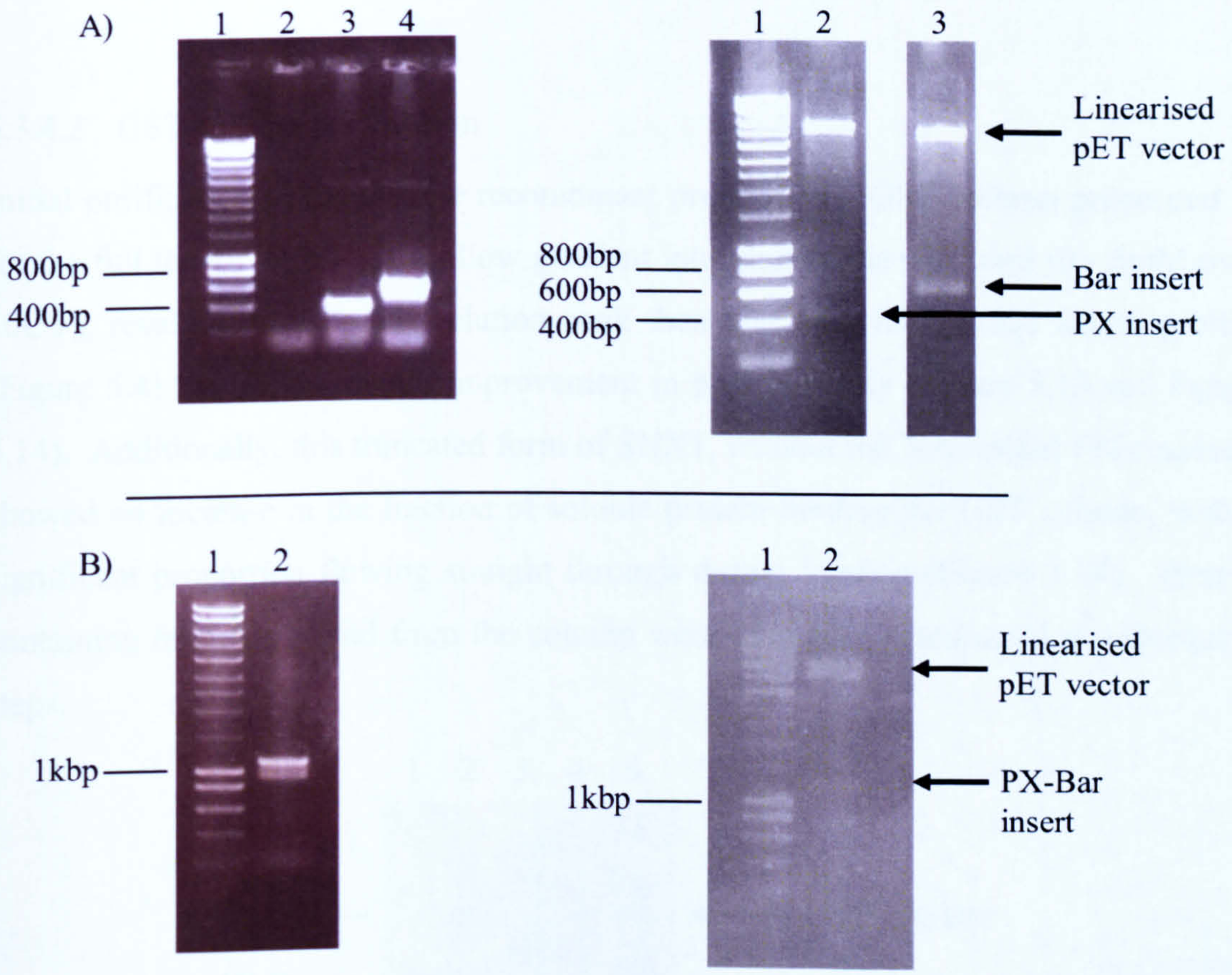


Figure 5.11 Agarose gels of poly-His tagged constructs.

1% agarose gels of; Lane 1) DNA ladder with relevant sizes indicated, A) PCR products of PX (3) and Bar (4) domain coding sequence, with final pET vector conformation by double digestion in the second gel (PX: 2, Bar: 3). B) PCR product of PX-Bar (2, care was taken to only take the upper band), with final pET vector conformation by double digestion in the second gel. All show the correct sized insert of approximately 400pb (PX), 600bp (Bar) and 1.1kbp (PX-Bar). Linearised pET vector is approximately 5.4kpb. All inserts were also confirmed by DNA sequencing.

5.3.4 GST-PX-Bar

5.3.4.1 Expression test

To investigate the effect of the removal of the unstructured N-terminal region of SNX1 on soluble expression, small scale expressions were performed at 16°C and 37°C with two different IPTG induction concentrations (0.1mM and 1mM). Soluble expression, however, was only achieved with 0.1mM ITPG induction and expression at 16°C (Figure 5.12), the same conditions as for full length GST-SNX1. Large scale expressions were therefore undertaken as described previously (5.2.1.1).

5.3.4.2 GST column purification

Initial purification of the soluble recombinant protein via a GST column proceeded as for the full length SNX1. A shallow gradient into glutathione was used (0 - 5mM over 10CV), resulting in a broader elution peak then that seen for the full length protein (Figure 5.4) but no significant improvement in protein purity (Figure 5.13 and Figure 5.14). Additionally, this truncated form of SNX1, without the N-terminal 138 residues, showed no increase in the fraction of soluble protein binding the GST column, with a significant proportion flowing straight through during loading (Figure 5.14). Protein containing fractions eluted from the column were then pooled and used in subsequent steps.

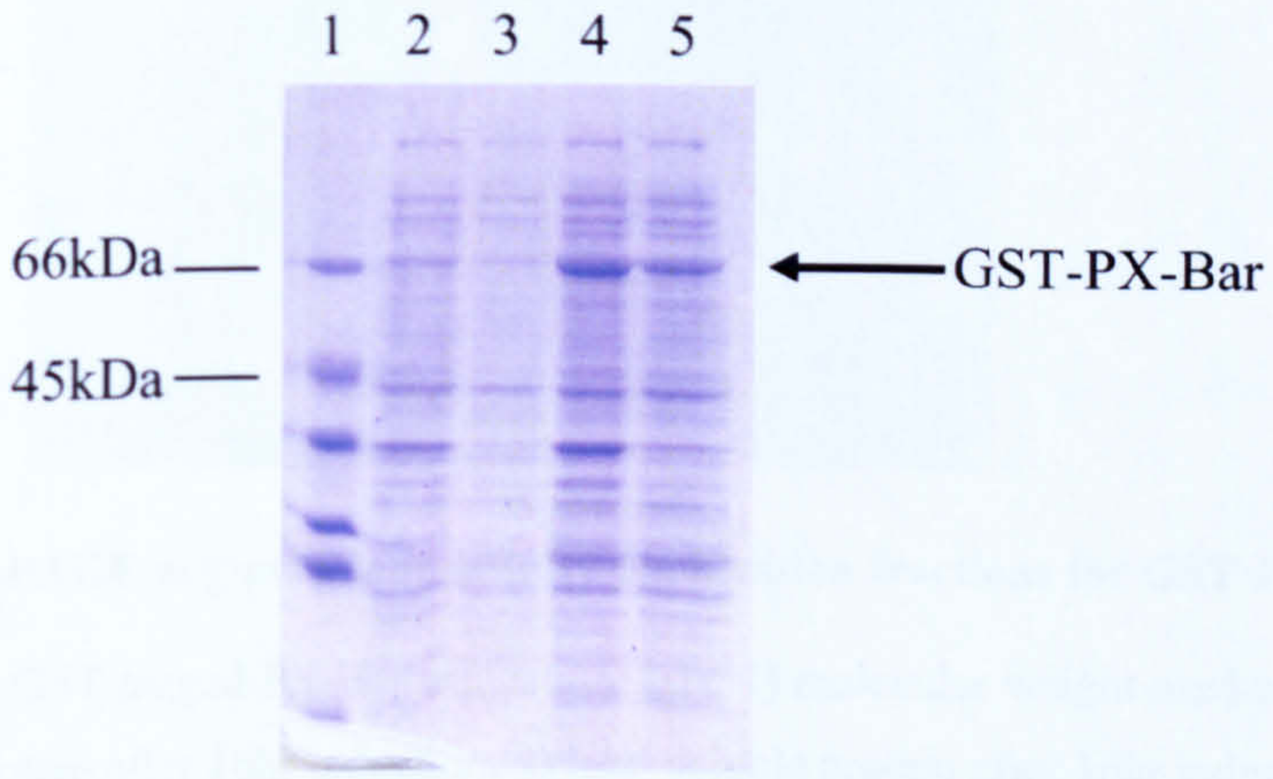


Figure 5.12 SDS-PAGE gel of 16°C expression of GST-PX-Bar.

Expression of the GST tagged PX-Bar construct undertaken at 16°C; lane 1) molecular weight markers, 2) uninduced total protein, 3) uninduced soluble protein, 4) total protein after overnight induction with 0.1mM IPTG, 5) soluble protein after overnight induction. GST-PX-Bar is produced at around its expected molecular weight (71kDa).

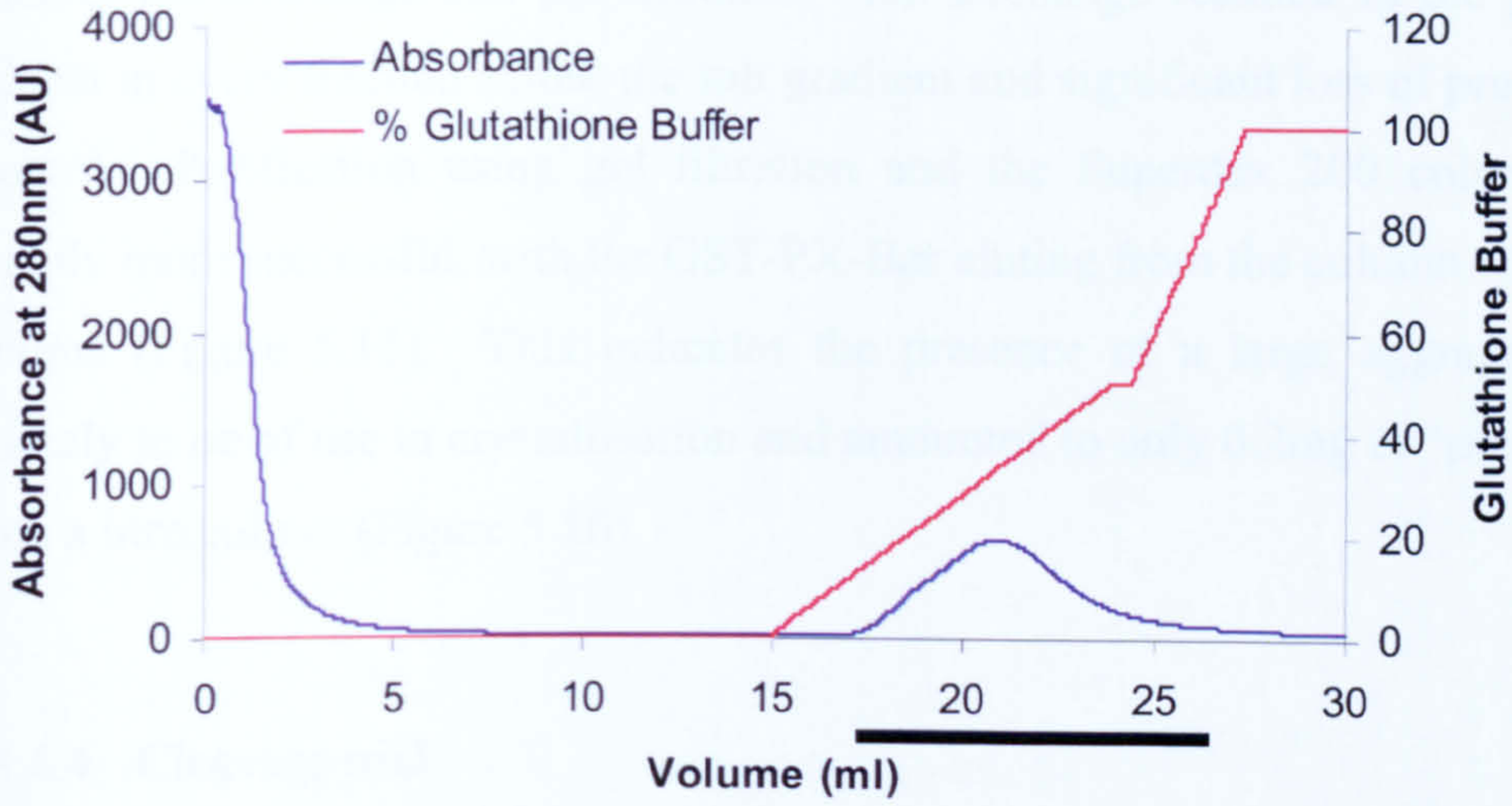


Figure 5.13 Glutathione affinity column chromatography of SNX1 PX-Bar truncation.

Profile of GST-PX-Bar elution run with a glutathione gradient, monitored by UV absorbance at 280nm. The solid line indicates those fractions ran on the SDS-PAGE gel in Figure 5.14.

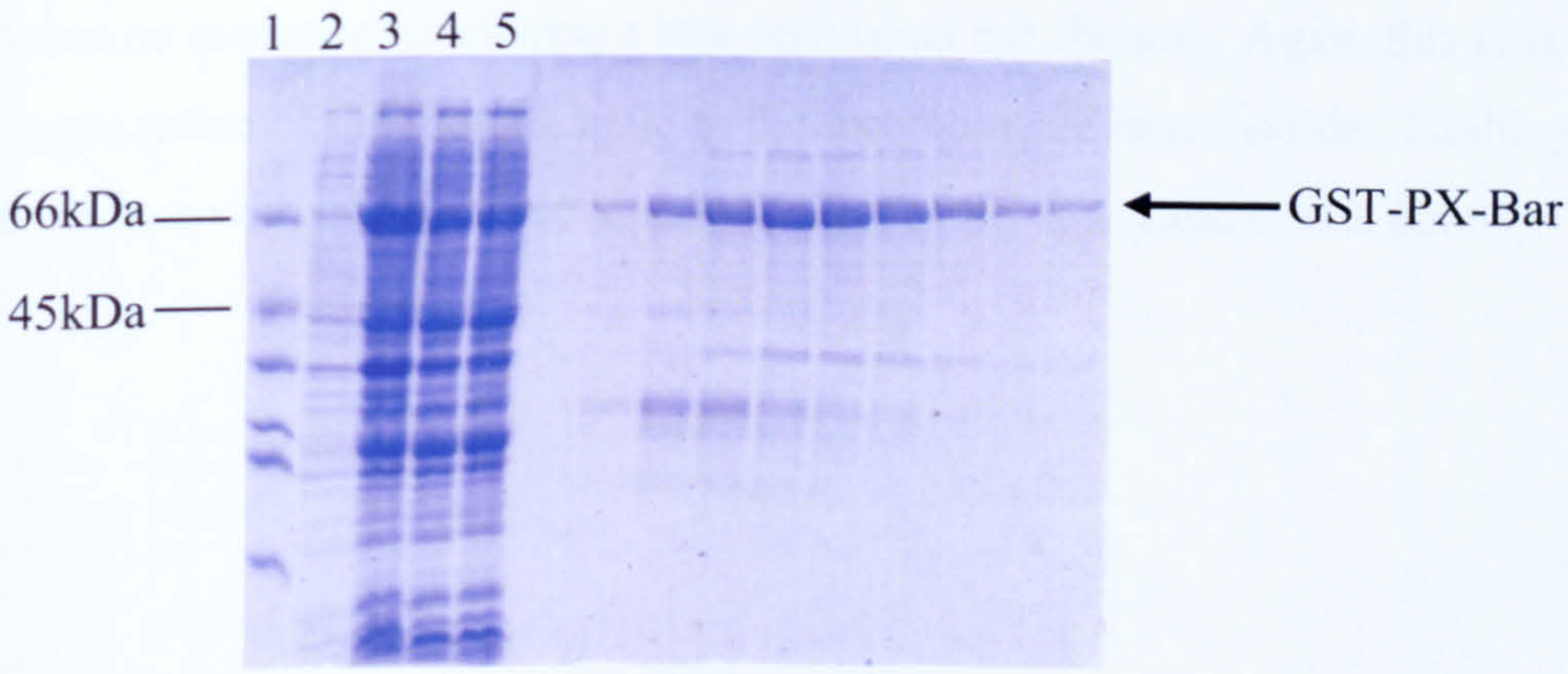


Figure 5.14 SDS-PAGE of glutathione affinity purification fractions for GST-PX-Bar.

Purification of the GST tagged PX-Bar truncation; lane 1) molecular weight markers, 2) uninduced total protein, 3) total protein after 16hr induction, 4) total soluble protein after 16hr induction, 5) flow through fraction from glutathione column chromatography. Other lanes correspond to fractions marked with a solid line in Figure 5.13.

5.3.4.3 Further purification attempts

Subsequent purification of the uncleaved GST-PX-Bar sample was attempted via MonoQ ion exchange and gel filtration. Ion exchange resulted in the protein being present in every fraction across the salt gradient and significant loss of protein (data not shown). Purification using gel filtration and the Superdex 200 column was only slightly more successful, with the GST-PX-Bar eluting from the column within the void volume (Figure 5.15). This indicates the presence of a large aggregate, which is unlikely to be of use in crystallisation and amounted to only 0.2mg of 'purified' protein from a litre culture (Figure 5.16).

5.3.4.4 Cleavage trial

A time course thrombin cleavage trial was undertaken on the glutathione column purified sample at 18°C. This resulted in a number of non-specific cleavage events and significant loss of protein, indicating some level of instability had been introduced with the removal of the N-terminal 138 amino acids (Figure 5.17). Reducing the temperature at which cleavage was undertaken did not improve this. Thrombin cleavage was also attempted on the sample from the previous gel filtration (Figure 5.15), but no evidence for cleavage was seen (data not shown). Again this is consistent with aggregation of the protein, making the thrombin site inaccessible. Nothing further could be done with this construct due to the extremely low yield at this point.

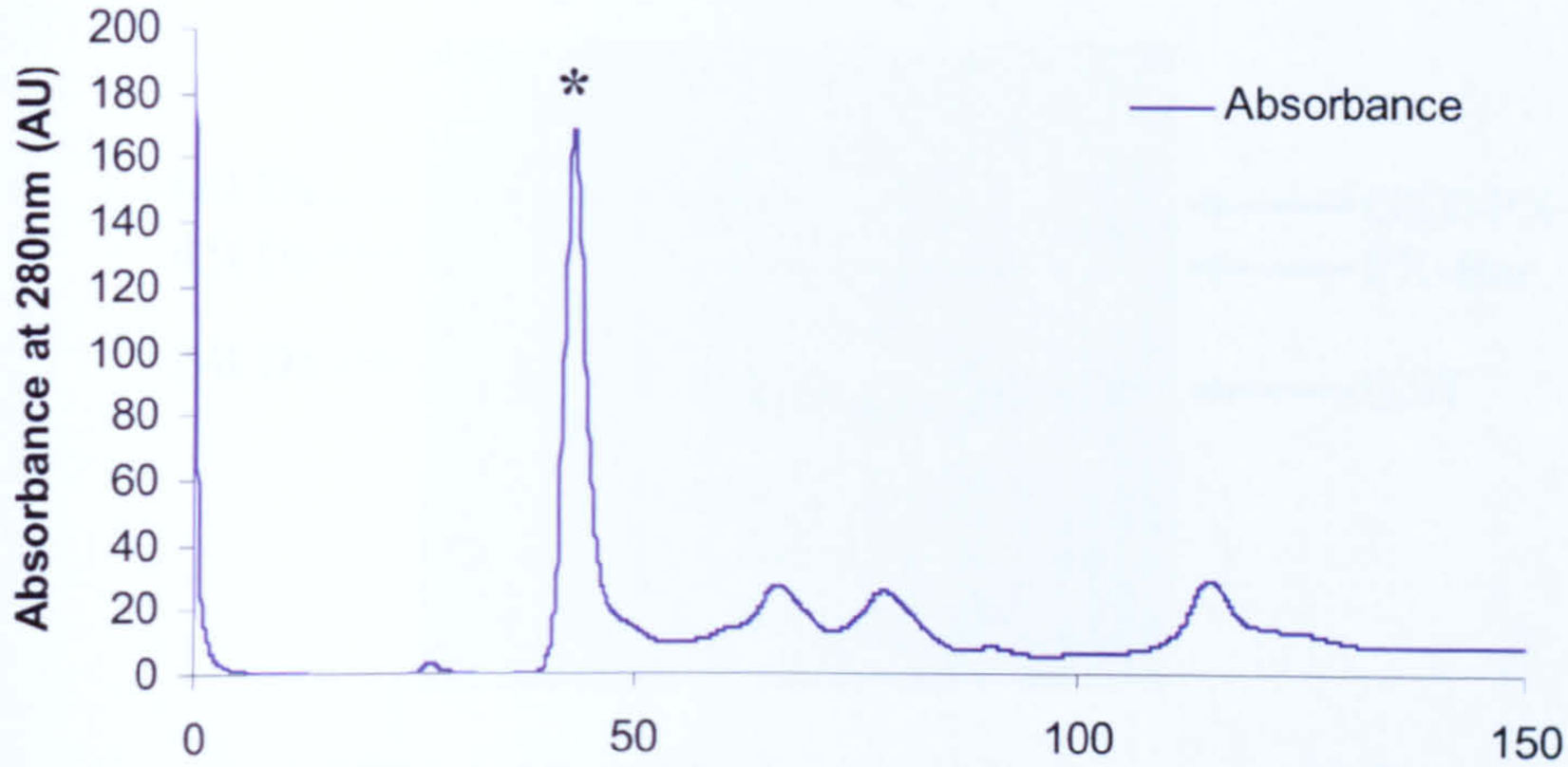


Figure 5.15 S200 gel filtration chromatography of GST-PX-Bar.

Protein elution was monitored by UV absorbance at 280nm. The only peak to contain GST-PX-Bar is indicated by the star. This corresponds to the void volume of the column indicating aggregate protein.

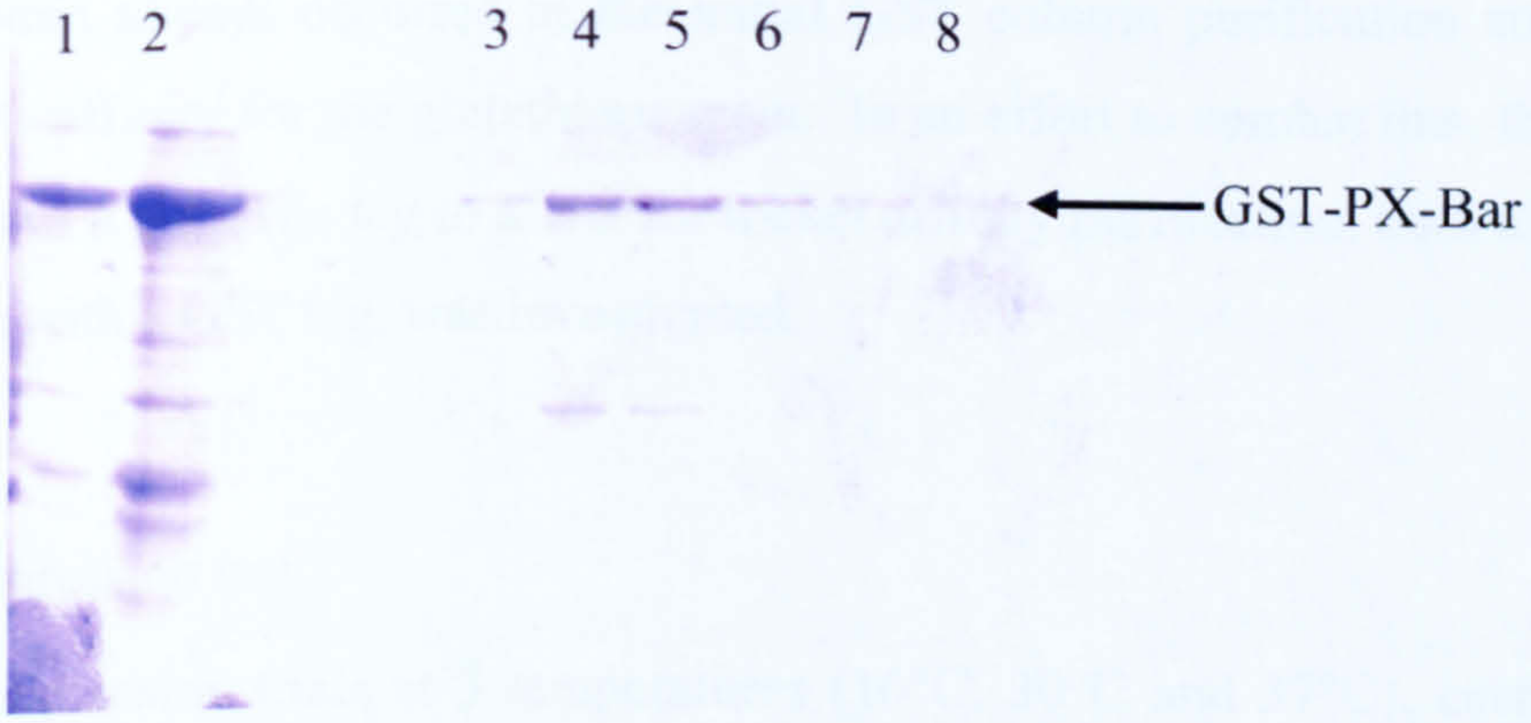


Figure 5.16 SDS-PAGE gel of aggregated GST-PX-Bar.

SDS-PAGE gel of S200 gel filtration chromatography of the GST tagged PX-Bar truncation; lane 1) pooled GST-PX-Bar from Figure 5.14, 2) concentrated GST-PX-Bar prior to gel filtration, 3 – 8) fractions corresponding to the peak indicated in Figure 5.15.

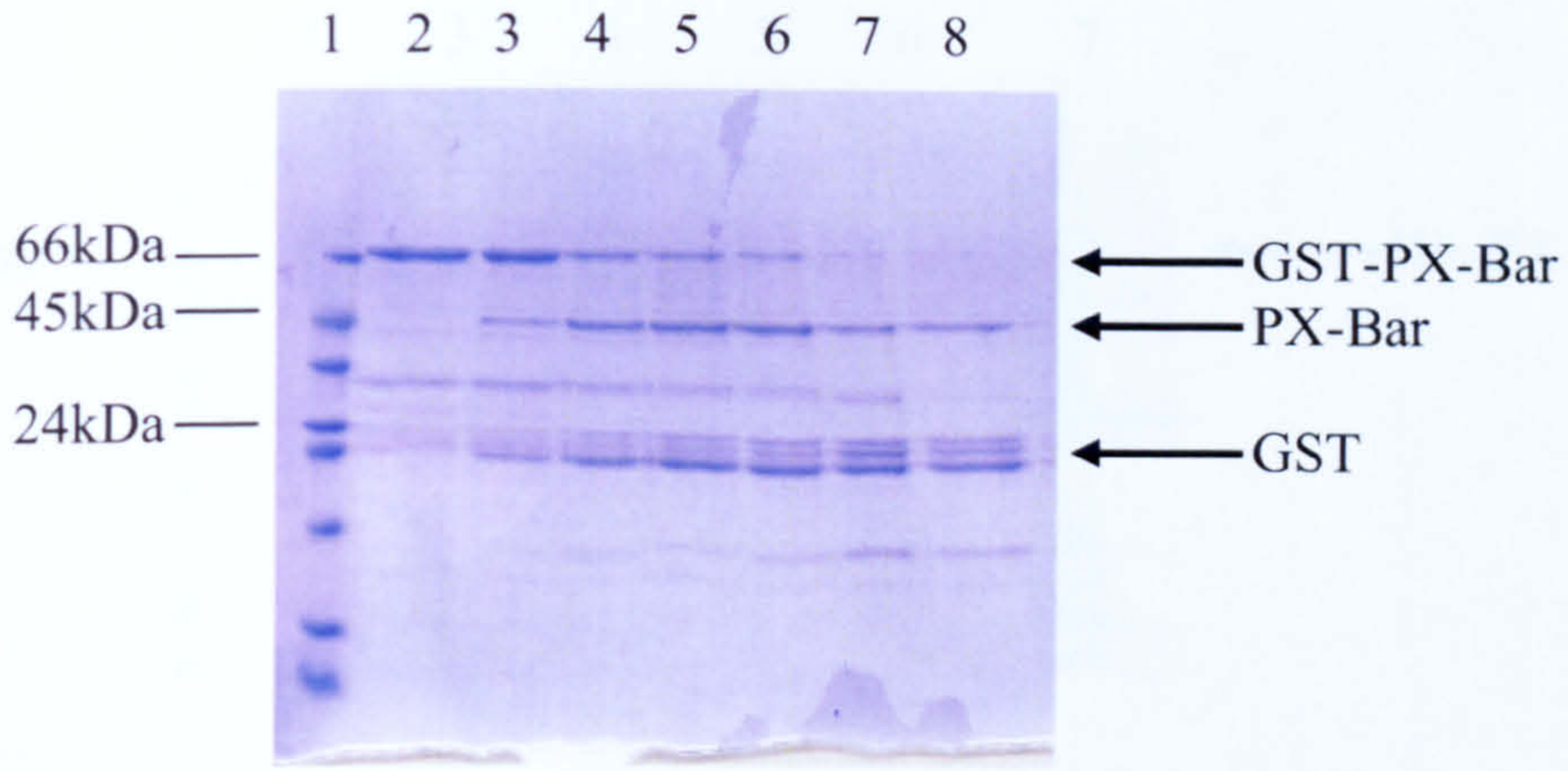


Figure 5.17 SDS-PAGE gel of thrombin cleavage trial undertaken on GST-PX-Bar.

Thrombin cleavage used to remove the GST tag from the PX-Bar SNX1 truncation purified in Figure 5.14; lane 1) molecular weight markers, 2) GST-PX-Bar pre-thrombin addition, 3) 2hrs post thrombin addition, 4) 4hrs, 5) 6hrs, 6) 8hrs, 7) 15hrs, 8) 23hrs Major cleavage products are labelled.

5.3.5 GST-PX-Bar-His and His-PX-Bar

With protein from previous constructs (GST-SNX1 and GST-PX-Bar) a large loss of soluble protein always occurred at the initial GST column purification stage due to apparent low affinity for the glutathione resin. In an effort to combat this, the effect of the addition of a poly-His tag to allow for nickel affinity purification, both alone and in conjunction with a GST tag, was investigated.

5.3.5.1 Expression test

Extensive expression trials at 3 temperatures (16°C, 30°C and 37°C), each with two IPTG concentrations (0.1mM and 1mM), were undertaken on both the GST-PX-Bar-His and His-PX-Bar constructs. Figure 5.18 shows an example expression trial for His-PX-Bar induced with 0.1mM IPTG at 16°C. At no point was soluble protein seen under any condition tested, nor was it during GST-PX-Bar-His expression (data not shown). This suggested that not only was the GST tag essential for soluble expression of the PX-Bar truncation, but that the addition of the hexa-His tag at the C-terminus was enough to destabilise this effect.

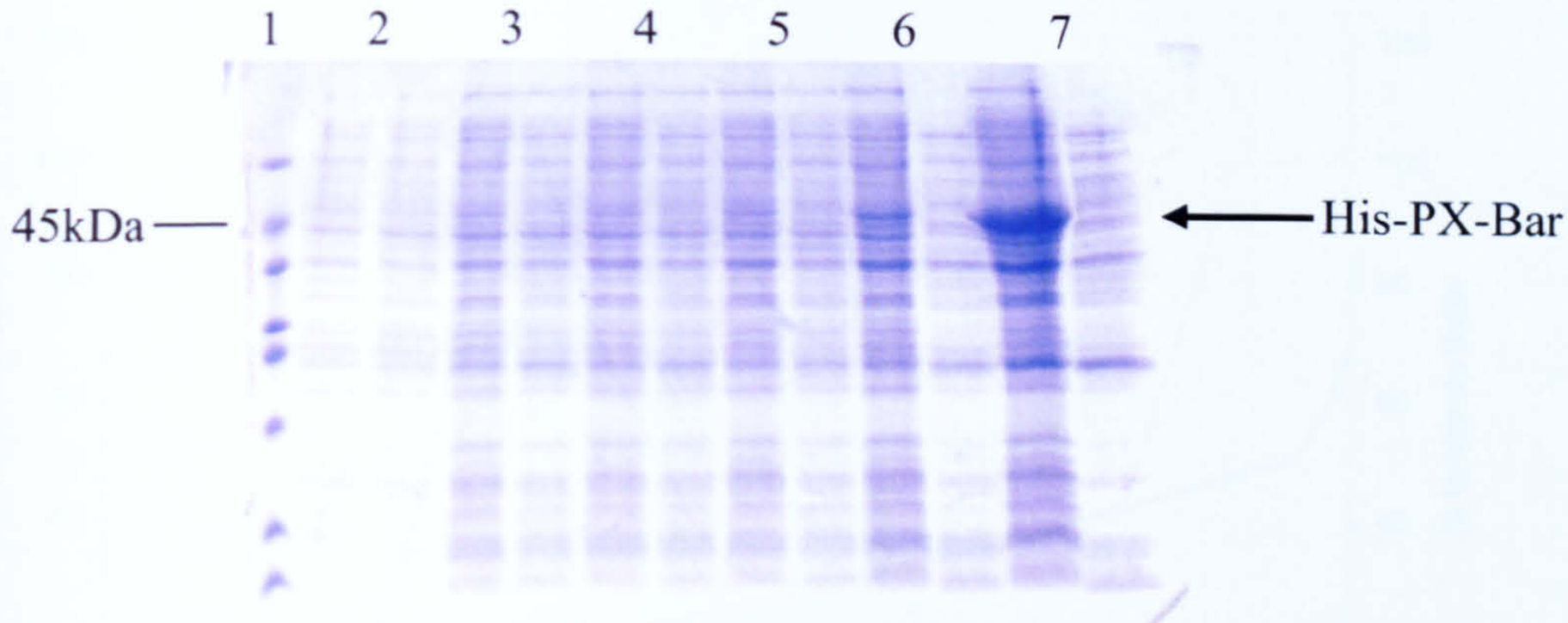


Figure 5.18 SDS-PAGE gel of expression trial for His-PX-Bar at 16°C.

Expression of the poly-His tagged PX-Bar construct undertaken at 16°C; lane 1) molecular weight markers. Each subsequent pair of lanes contains total protein followed by soluble protein at various time points after induction with 0.1mM IPTG; 2) uninduced, 3) 2hrs, 4) 4hrs, 5) 6hrs, 6) 8hrs, 7) 16hrs. Insoluble His-PX-Bar is seen appearing after 6hrs.

5.3.5.2 Denatured purification

In an effort to purify protein and allow refolding to be attempted, a denaturing Ni^{2+} column purification strategy was undertaken. Denatured His-tagged proteins can be easily purified, as their affinity for the Ni^{2+} resin is not dependent on a specific 3-dimensional structure. Denatured purification and subsequent refolding was only attempted on the GST-PX-Bar-His recombinant protein, as this was deemed to be the most likely to successfully refold.

Cell pellets from a GST-PX-Bar-His expression were treated as described in the methods, with the insoluble protein being resuspended in nickel column loading buffer containing 8M urea. This denatured protein was loaded onto an appropriately equilibrated nickel affinity column and unbound protein washed off. Elution of the bound, denatured protein from the column with a shallow imidazole gradient produced a reasonable A_{280} peak (Figure 5.19). Inspection of the resulting SDS-PAGE gel (Figure 5.20) showed GST-PX-Bar-His to be the major constituent of the peak, which eluted with approximately 250mM imidazole. Additionally, little recombinant protein was present in the flow through fraction and the resulting eluted protein was considered pure enough to proceed with refolding experiments.

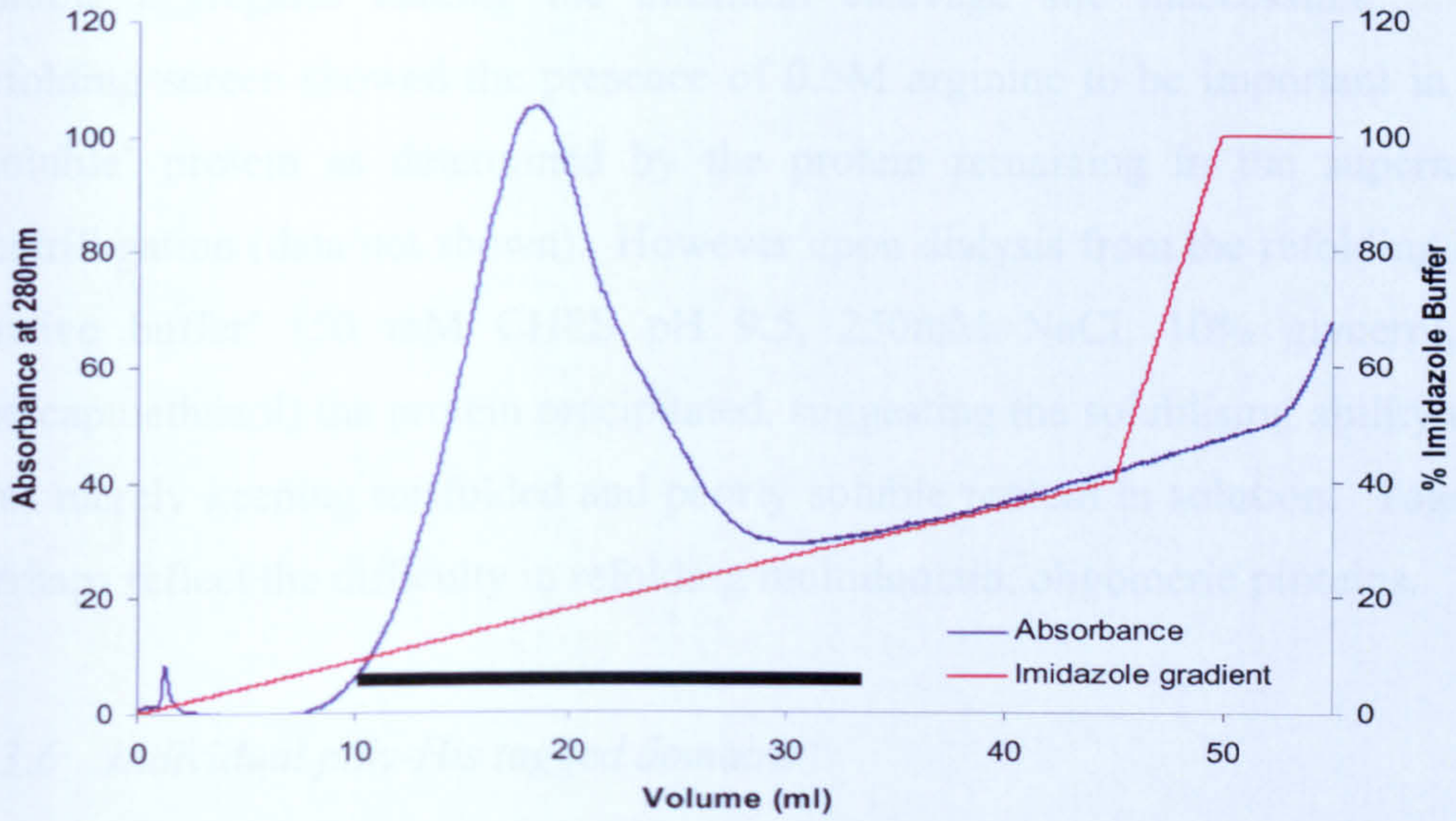


Figure 5.19 Denatured nickel affinity column chromatography of GST-PX-Bar-His.

The purification was carried out in the presence of 8M urea and monitored by UV absorbance at 280nm. Solid line indicates the range of fractions seen in Figure 5.20

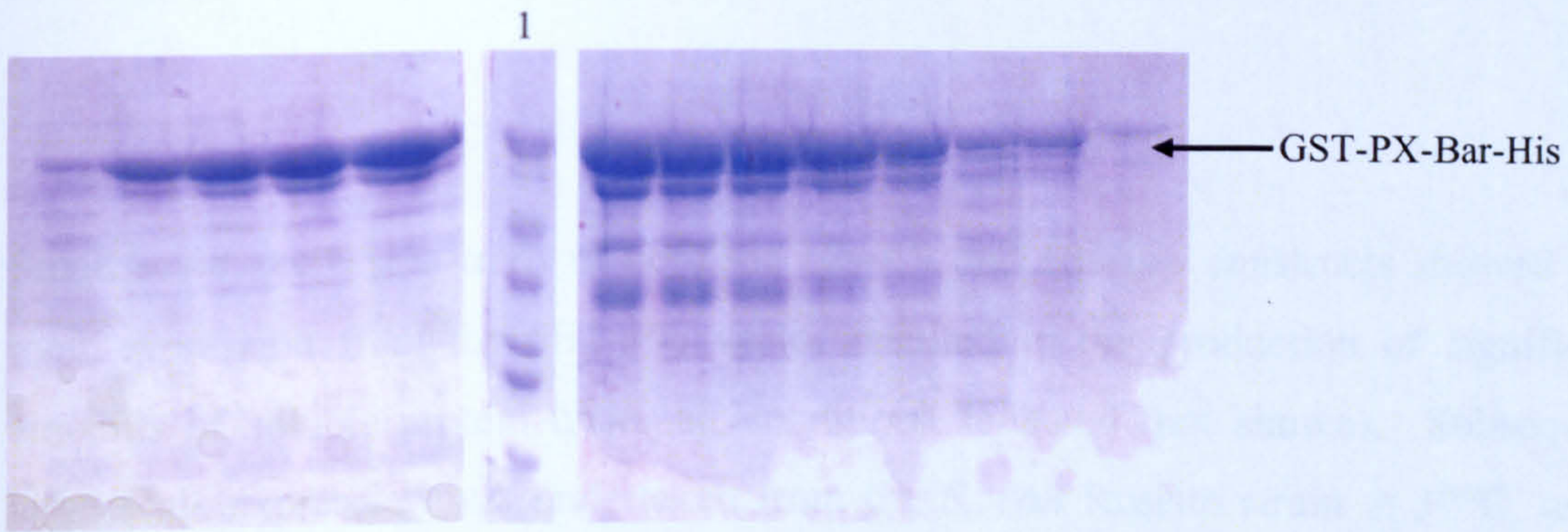


Figure 5.20 SDS-PAGE of the denatured nickel affinity purification of GST-PX-Bar-His.

Denatured purification of the dual tagged PX-Bar truncation; 1) molecular weight markers. Other lanes (from left to right) correspond to fractions marked with a solid line in Figure 5.19.

5.3.5.3 Refolding

Various refolding techniques such as rapid dilution and dialysis were attempted as described in 5.2.6. Both the MDL screen, rapid dilution and dialysis buttons gave initially encouraging results. Refolding by dialysis into 50 mM CHES pH 9.5 appeared to produce soluble protein (as determined by centrifugation), but the addition of thrombin did not result in any cleavage. This suggested misfolding or the formation of

soluble aggregates making the thrombin cleavage site inaccessible. The MDL refolding screen showed the presence of 0.5M arginine to be important in producing 'soluble' protein as determined by the protein remaining in the supernatant after centrifugation (data not shown). However upon dialysis from the refolding buffer into 'native buffer' (50 mM CHES pH 9.5, 250mM NaCl, 10% glycerol, 1mM β-mercaptoethanol) the protein precipitated, suggesting the solublising ability of arginine was merely keeping misfolded and poorly soluble protein in solution. Together these perhaps reflect the difficulty in refolding multidomain, oligomeric proteins.

5.3.6 Individual poly-His tagged domains

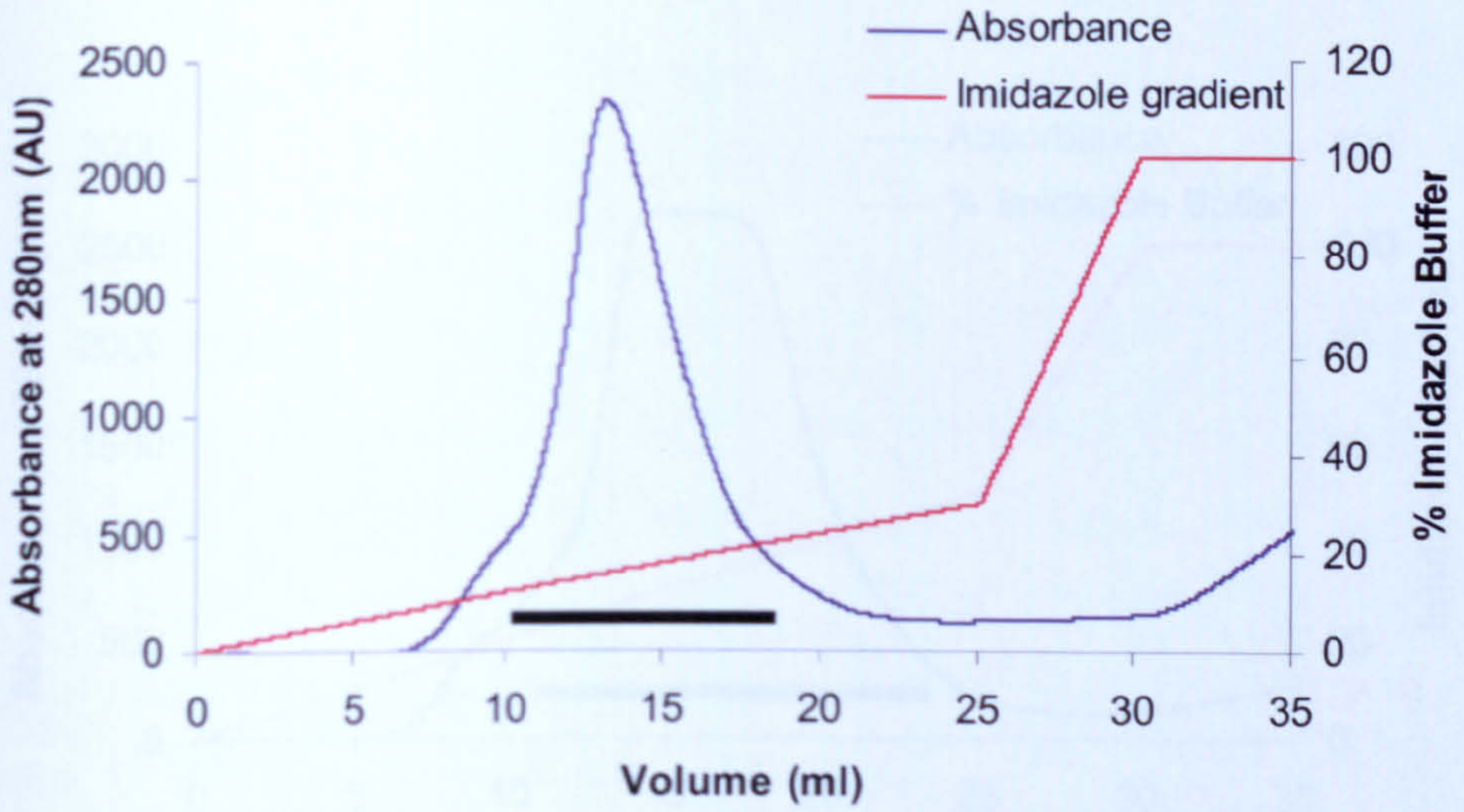
Although it had not proven feasible to purify full-length SNX1 in adequate quantities, the discrete domains that make up the SNX1 protein are thought to act in a separate, but additive, manner to control the specificity of the overall protein (Carlton *et al.*, 2004). The structures of the isolated domains should therefore provide a lot of the specificity information, while leaving only their spatial relationship unclear.

5.3.6.1 Expression

Small scale expression trials of both the His-PX and His-Bar constructs showed that their expression from the pET-30b vector resulted in the production of significant amounts of soluble protein under all conditions screened (not shown). Subsequent large scale expression was undertaken from the *E. coli* Rosetta strain at 37°C, using 1mM IPTG to induce expression for 3 - 5 hrs as these conditions produced the best yield.

5.3.6.2 Nickel column purification

Purification of the expressed domains proceeded via nickel affinity chromatography using standard procedures. This resulted in a large increase in protein purity for both the His-PX and His-Bar proteins. The His-PX protein eluted from the Ni^{2+} column with approximately 150mM imidazole (Figure 5.21) and the elution of a large His-Bar A_{280} peak correlated to 300mM imidazole (Figure 5.22). Contaminating proteins present in these fractions were far outweighed by the amount of recombinant protein and tag cleavage was undertaken directly on pooled samples.



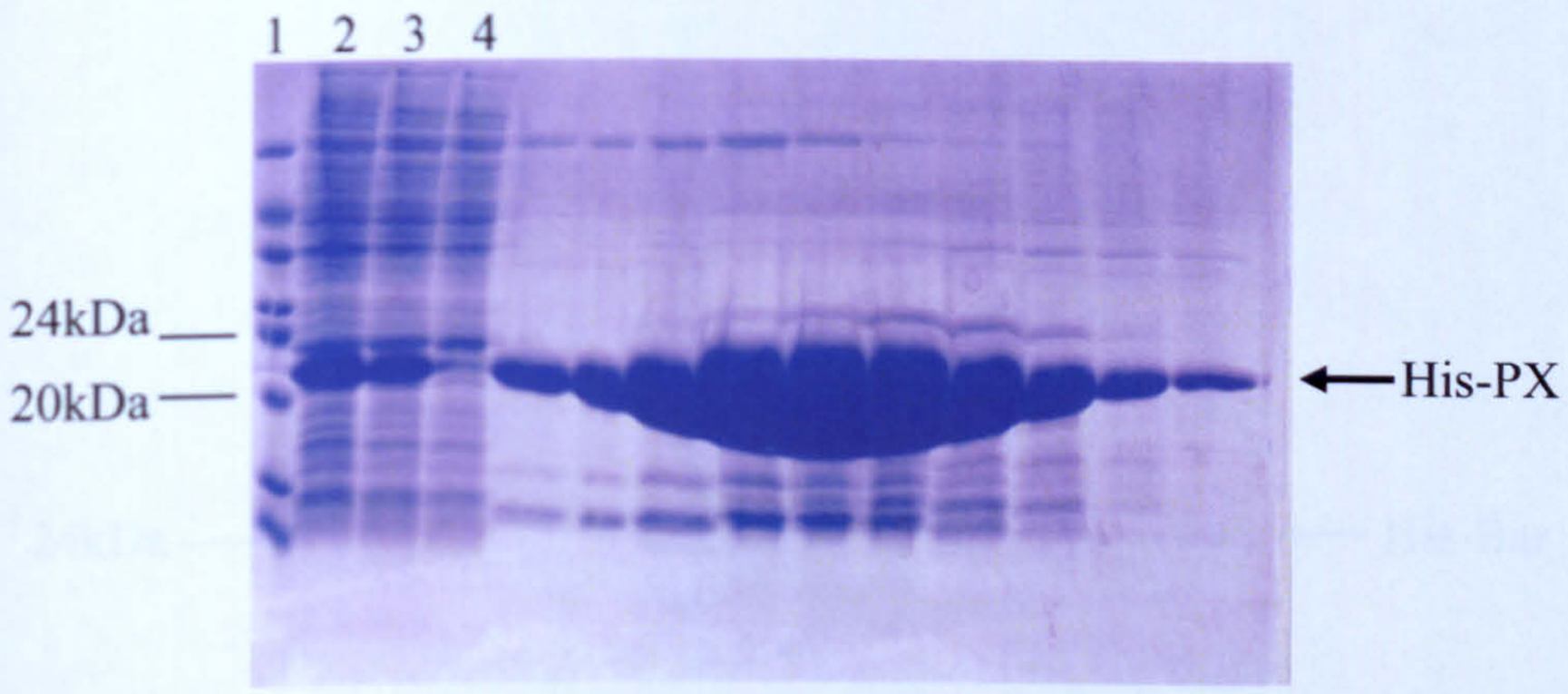
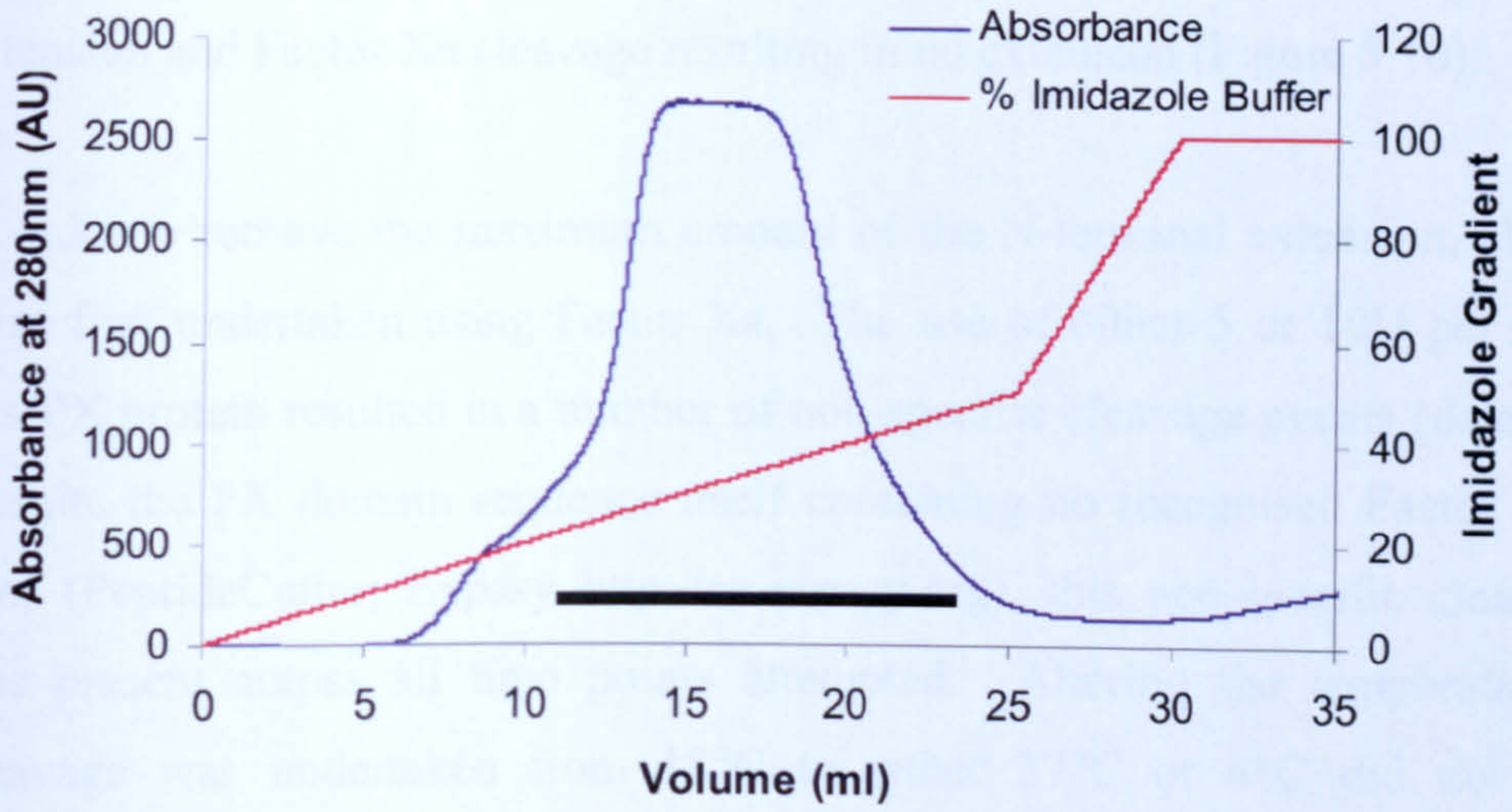


Figure 5.21 Nickel affinity purification of His-PX.

Chromatography trace and resulting SDS-PAGE gel of the nickel affinity purification of His-PX. Protein elution from the nickel column was monitored by UV absorbance at 280nm. Within the SDS-PAGE gel lanes correspond to; 1) molecular weight marker, 2) total protein, 3) total soluble protein, 4) flow-through from column loading. The remaining lanes correspond to the solid black line present on the chromatography trace.



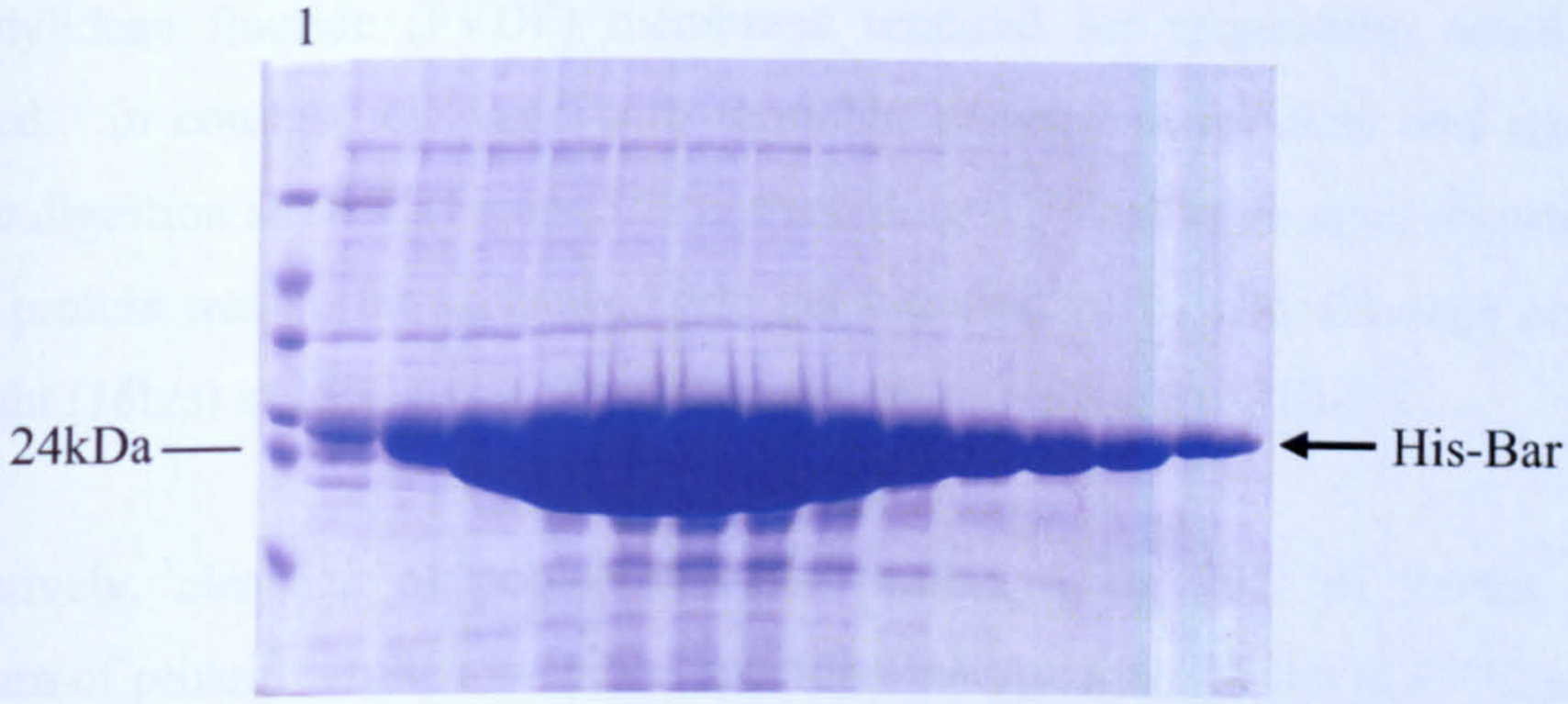


Figure 5.22 Nickel affinity purification of His-Bar.

Chromatography trace and resulting SDS-PAGE gel of the nickel affinity purification of His-Bar. Protein elution from the nickel column was monitored by UV absorbance at 280nm. Within the SDS-PAGE gel lanes correspond to; 1) molecular weight marker, and the remaining lanes to the solid black line present on the chromatography trace.

5.3.6.3 His-tag removal

Due to the construction of the recombinant protein the His-tag could theoretically be removed by the use of either thrombin or Factor Xa, with thrombin leaving a 24aa extension and Factor Xa cleavage resulting in no extension (Figure 5.10).

In order to remove the maximum amount of the N-terminal extension, cleavage trials were first undertaken using Factor Xa. The use of either 5 or 10U per milligram of His-PX protein resulted in a number of non-specific cleavage events (data not shown). Despite the PX domain sequence itself containing no recognised Factor Xa cleavage sites (PeptideCutter; Expasy http://ca.expasy.org), this non-specific cleavage pattern was present across all time points attempted. Altering the temperature at which cleavage was undertaken from 18°C to either 37°C or 4°C did not modify the apparently poor specificity, with the same cleavage pattern occurring (data not shown). Efforts were made to N-terminally sequence the aberrant cleavage products to clarify where they were occurring. However, despite many attempts, transfer to the Polyvinylidene fluoride (PVDF) membrane required for sequencing could not be achieved. In contrast, cleavage with thrombin resulted in efficient and reasonably specific digestion at 18°C (Figure 5.23). Preceding the final large scale digestions, the pooled protein was buffer exchanged into gel filtration buffer and cleavage performed overnight (16hrs) at 18°C using 5U of thrombin per milligram of His-PX.

Alternatively, cleavage of pooled His-Bar protein using 10U of Factor Xa per milligram of protein resulted in stable, specific cleavage after 16 hrs at 18°C, although a low molecular weight doublet can be seen appearing across the course of digestion (Figure 5.24). Again, preceding large scale digestions the protein was buffer exchanged into the appropriate gel filtration buffer.

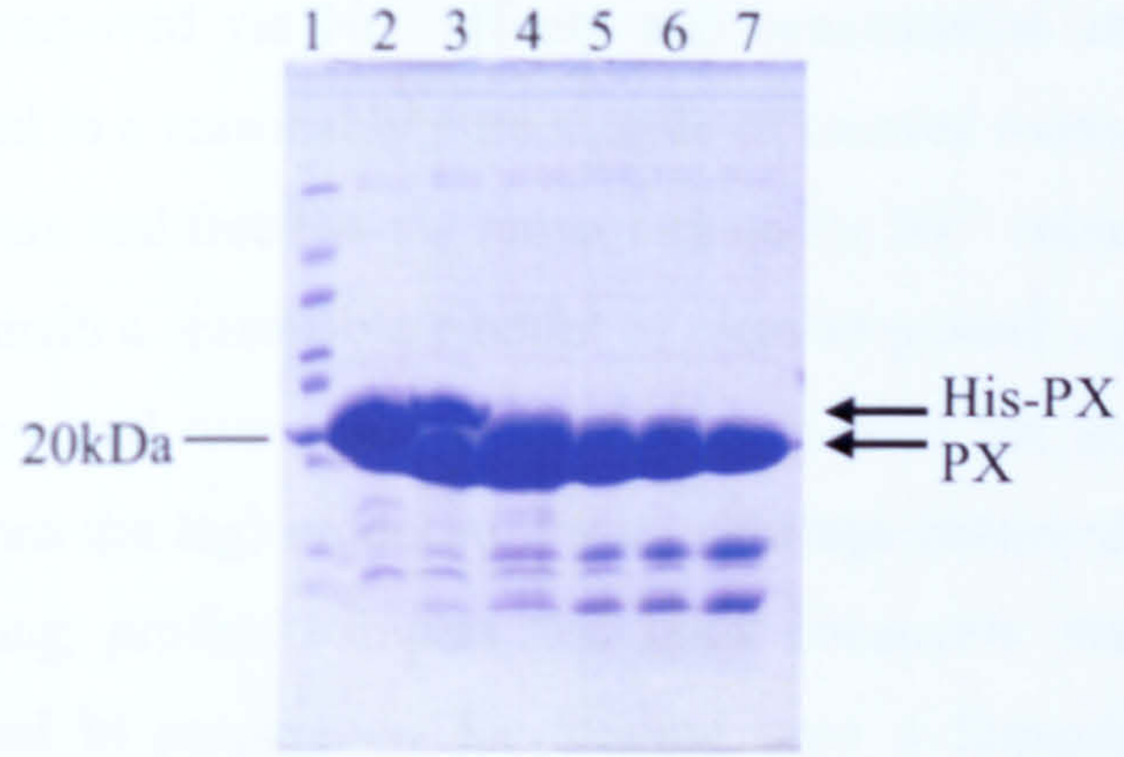


Figure 5.23 Thrombin cleavage of His-PX.

SDS-PAGE gel of time course digestion of His-PX with 5U thrombin per mg to remove the poly-His tag. Lane 1) molecular weight marker, 2) uncleaved His-PX, 3) 1hr post thrombin addition, 4) 2hrs, 5) 3hrs, 6) 8hrs, 7) 16hrs. Major cleavage products are labelled.

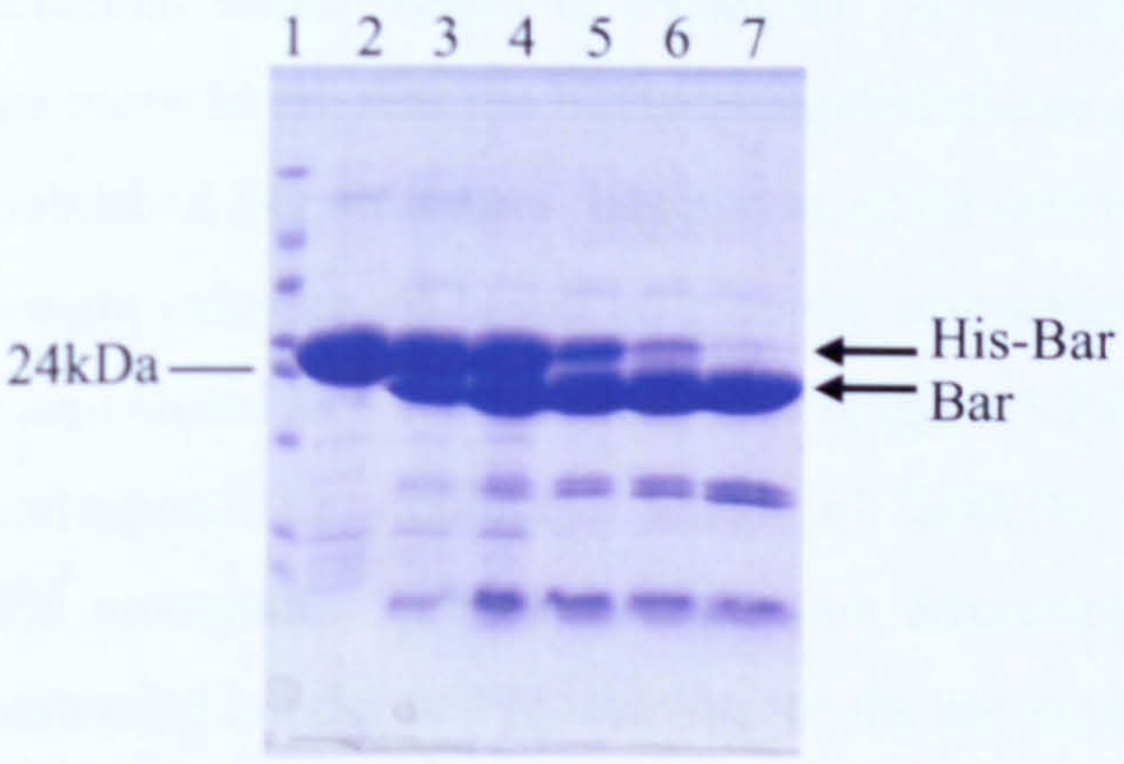


Figure 5.24 Factor Xa cleavage of His-Bar.

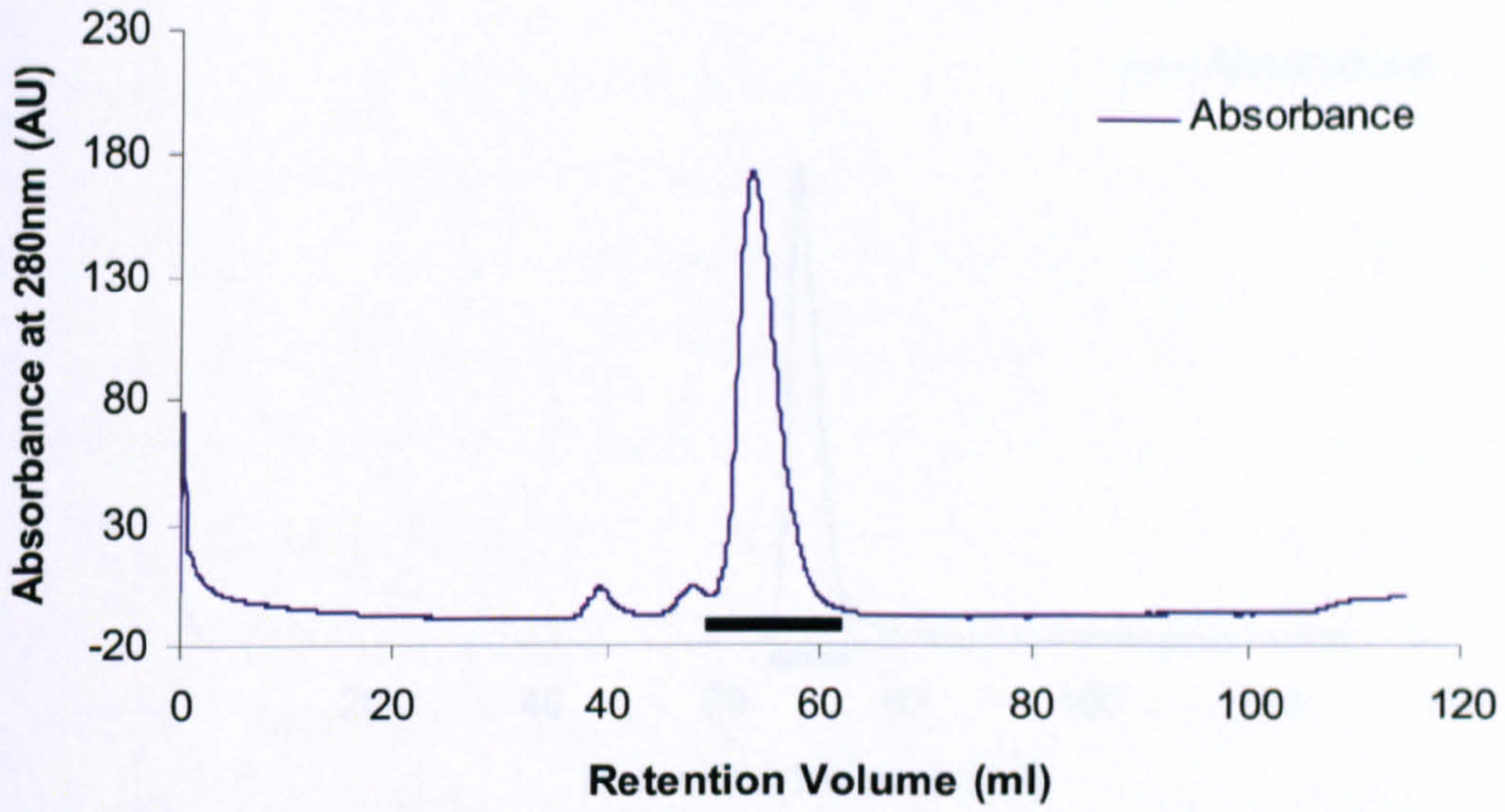
SDS-PAGE gel of time course digestion of His-Bar with 10U Factor Xa per mg to remove the poly-His tag. Lane 1) molecular weight marker, 2) uncleaved His-Bar, 3) 1hr post thrombin addition, 4) 2hrs, 5) 3hrs, 6) 8hrs, 7) 16hrs. Major cleavage products are labelled.

5.3.6.4 Gel Filtration

Prior to separation by gel filtration, any uncleaved protein and the serine protease used for cleavage was removed via Ni^{2+} affinity and benzamidine columns connected in series. This resulted in a reasonably pure sample of cleaved recombinant protein, with any uncleaved protein and free His-tag remaining on the Ni^{2+} column. However, in the case of the Bar domain a reasonable amount of cleaved protein remained bound to the benzamidine column and was lost. In neither case were the low molecular weight fragments (other then the tag) produced during cleavage removed by this purification step. The resulting protein samples for both constructs were concentrated to approximately 0.5ml in preparation for loading onto a Superdex 75 gel filtration column. The column was run slowly (0.3ml.min^{-1}) and smaller then normal fractions (1ml) were collected in order to achieve the degree of resolution required to separate the recombinant domains from the smaller contaminating proteins.

Purified Bar domain eluted from the column as a large symmetrical peak at a volume of 54.2ml (Figure 5.25). This corresponds to a molecular weight of 86.5kDa (logMr = -1.4758($V_{elution}/V_{viod}$) + 6.8879, where V_{viod} = 41), 1.75 times the size of the expect dimer (predicted molecular dimer weight of 49.4kDa). A dimer of dimers might be possible, but it seems more likely that the likely elongated nature of the Bar dimer, as expected based on published Bar structures, has produced an erroneous elution volume making molecular weight calculations based on globular standards inappropriate. The purified PX domain also eluted as a symmetrical peak (Figure 5.26). This peak elution volume of 68.5ml corresponds to a molecular weight of 26.4kDa, half as big again as the expect 18kDa PX monomer. The reasons for this discrepancy are unclear, but could be due to the extended tag derived N-terminus or the molecular shape. A similar size discrepancy has been seen previously with gel filtration of SNX1-PX (Zhong *et al.*, 2002) and was shown not to be due to any dimerisation.

After gel filtration and careful selection of fractions to be combined, approximately 4mg of each domain was produced and deemed to be of greater then 95% purity by SDS-PAGE. Pooled fractions were buffer exchanged into their respective crystallisation buffer (5.2.7.4). Crystallisation trials were generally attempted immediately using 'fresh' protein, but the domains seemed stable when stored at -20°C or 4°C for a short period.



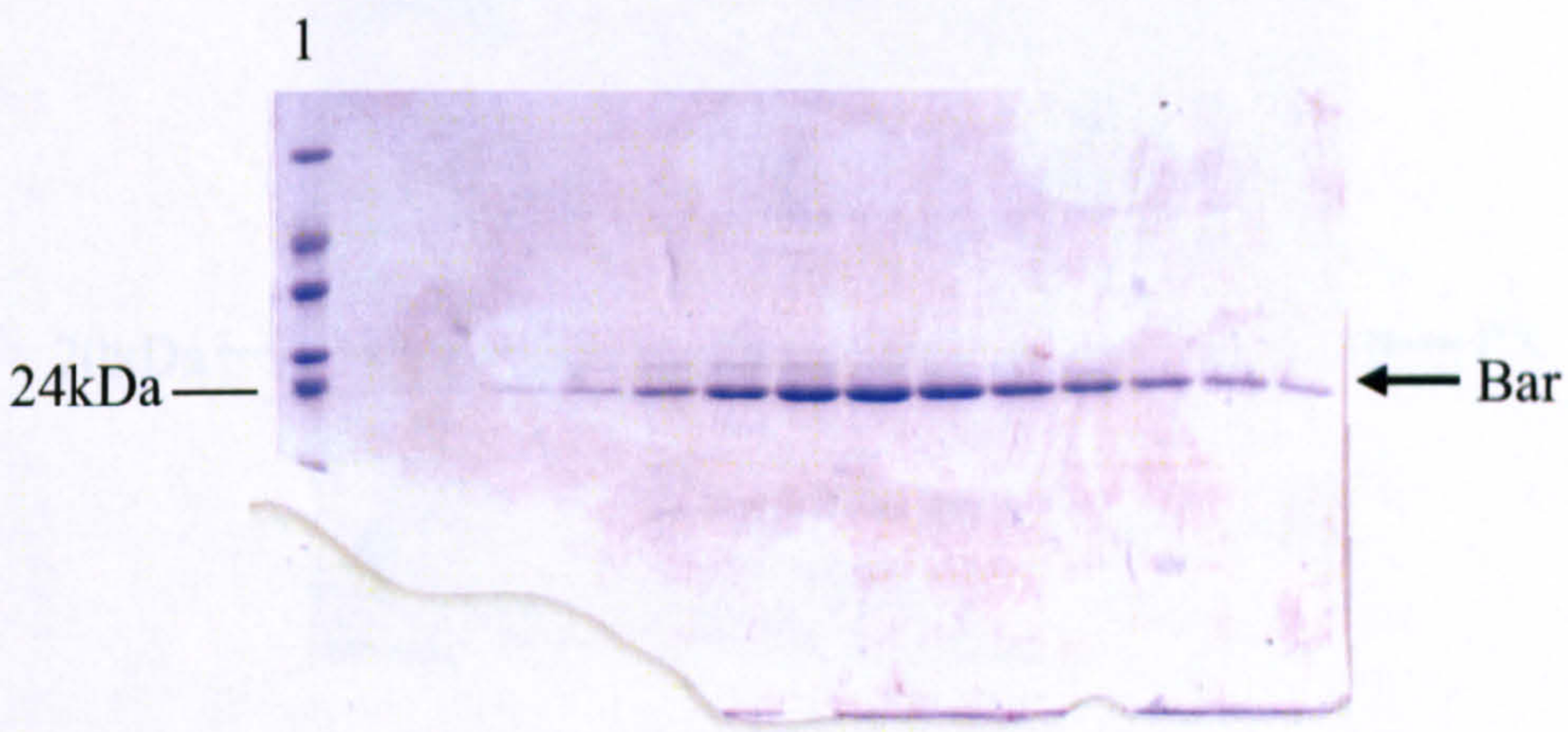
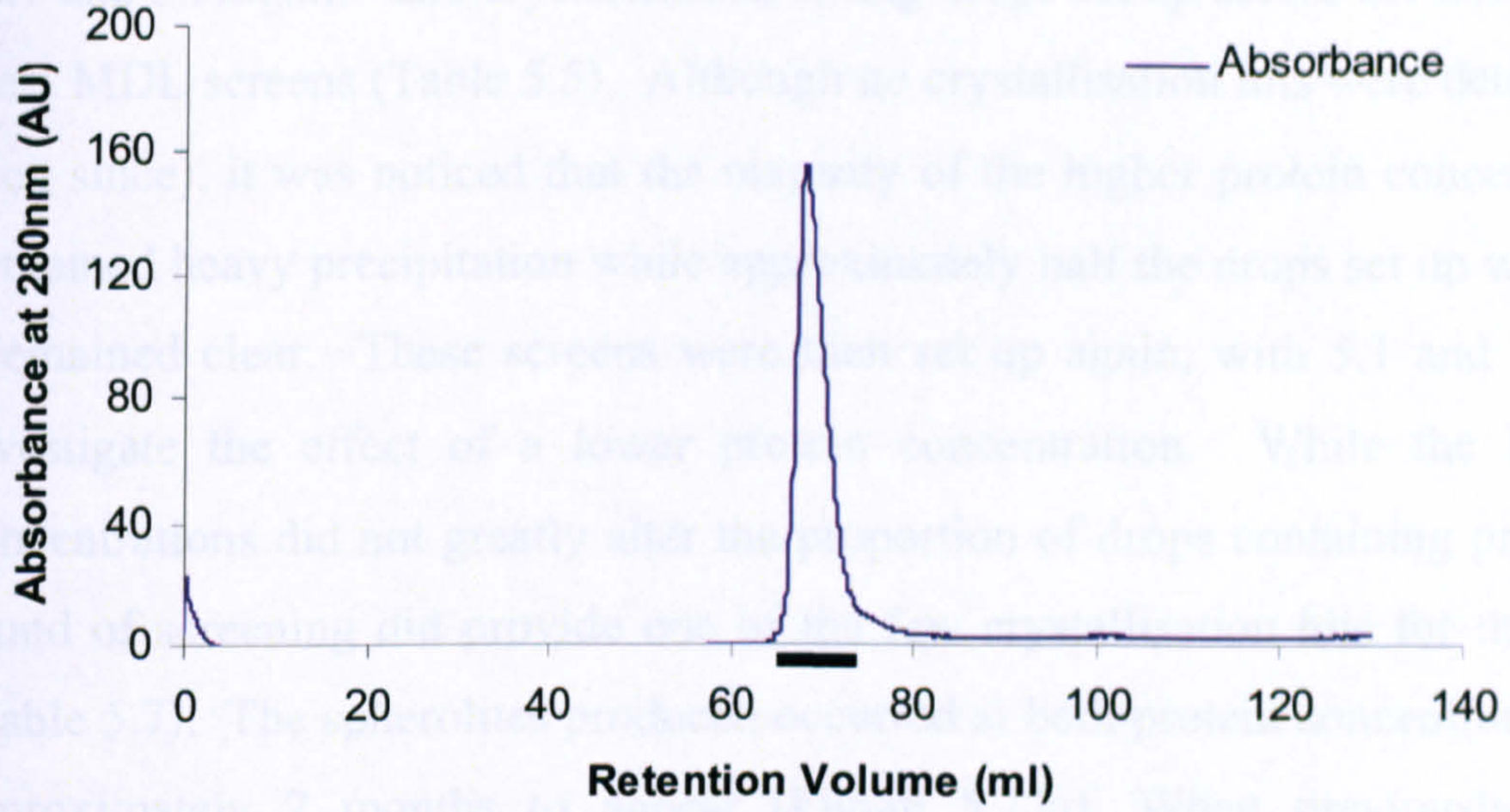


Figure 5.25 Gel filtration purification of cleaved Bar domain.

Chromatography trace and resulting SDS-PAGE gel of the gel filtration Factor Xa cleaved Bar domain. Protein elution from the size exclusion column was monitored by UV absorbance at 280nm. Within the SDS-PAGE gel lanes correspond to; 1) molecular weight marker, and the remaining lanes to the solid black line present on the chromatography trace.



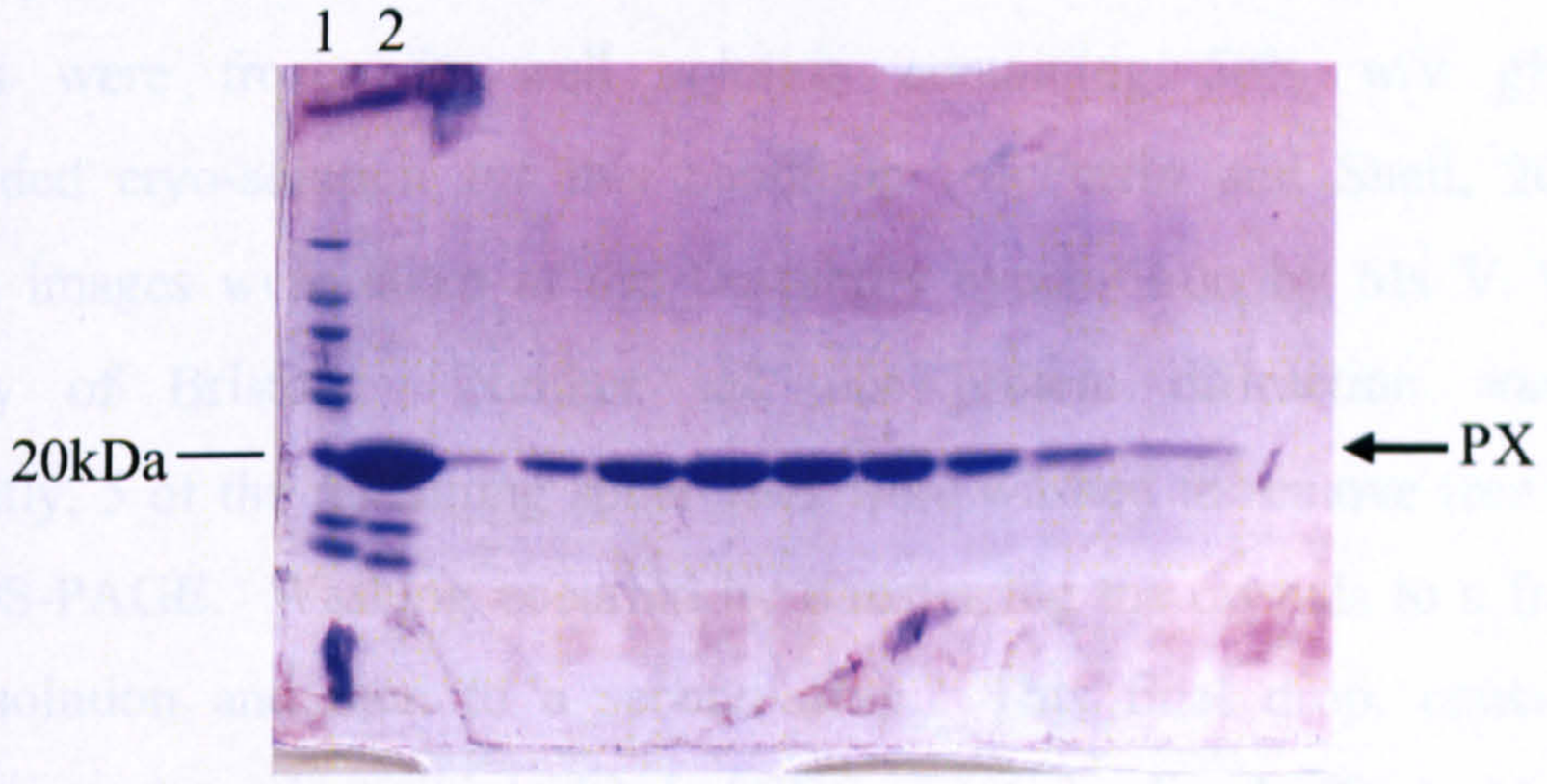


Figure 5.26 Gel filtration purification of cleaved PX domain.

Chromatography trace and resulting SDS-PAGE gel of the gel filtration thrombin cleaved PX domain. Protein elution from the size exclusion column was monitored by UV absorbance at 280nm. Within the SDS-PAGE gel lanes correspond to; 1) molecular weight marker, 2) pooled, cleaved PX domain prior to size exclusion chromatography. The remaining lanes correspond to the solid black line present on the chromatography trace.

5.3.6.5 Crystallisation

Extensive screening using commercially available crystallisation screens was undertaken with both purified domains. The PX domain was initially concentrated to 16.7 and 5.5mg.ml^{-1} and crystallisation sitting drops set up across the 288 conditions of the 3 MDL screens (Table 5.5). Although no crystallisation hits were detected (or have been since), it was noticed that the majority of the higher protein concentration drops contained heavy precipitation while approximately half the drops set up with 5.5mg.ml^{-1} remained clear. These screens were then set up again, with 5.1 and 3.4mg.ml^{-1} to investigate the effect of a lower protein concentration. While the lower protein concentrations did not greatly alter the proportion of drops containing precipitant, this round of screening did provide one of the few crystallisation hits for the PX domain (Table 5.7). The spherolites produced occurred at both protein concentrations and took approximately 2 months to appear (Figure 5.27a). When previously screened at 5.5mg.ml^{-1} this condition had remained clear, but it seems unlikely that the small drop in protein concentration is responsible for this and is more likely due to variation in the protein preparation. To investigate whether these were salt or protein crystals, the spherolites were frozen in well solution containing 50% w/v glycerol, the recommended cryo-solution for this condition (McFerrin and Snell, 2002). Test diffraction images were taken at the Daresbury synchrotron by Ms V. Fairweather (University of Bristol). Neither salt nor protein diffraction was detected. Subsequently, 5 of the remaining spherolites were washed to remove free protein and run on SDS-PAGE. Washing occurred by transferring the crystals to a fresh drop of reservoir solution and then to a second drop. This final drop, containing all 5 spherolites, was then pipetted into 10µl of SDS sample loading buffer and boiled for 15 minutes. The resulting Coomassie stained gel showed no protein present (not shown). If these crystals are protein rather then salt, this could be due to there being insufficient protein to detect by this method. Finally, the recent purchase of Izit Crystal Dye (Hampton Research) allowed this to be employed. This blue dye differentiates salt from protein crystals as the latter contains large solvent channels which Izit can penetrate, colouring the crystals a darker blue then the surrounding solvent. A 1:10 dilution of Izit was added (0.5µl) directly to a crystallisation drop containing some remaining spherolites. An hour later the crystals had become significantly darker then their surroundings, suggesting them to be protein crystals (Figure 5.27b).

Further crystallisation attempts were set up using the JBScreens. All 10 screens were set up using the Phoenix robot as sitting drops at 2.4 and 5.5mg.ml^{-1}, or a selection (JBScreen 3, 5, 6 and 7) were also done as hanging drops manually with 3mg.ml^{-1} of protein. From both methods, a total of 576 drops, only 3 drops produced a 'hit', each constituting phase separation (Table 5.7 and Figure 5.27c). A small amount of manual screening around the hits presented in Table 5.7 has been attempted as described in 2.3.15.2. However in no case were the hits reproducible. This could be due to the move from a robot set up sitting drop to a larger, manually set up hanging drop or simple variation in the protein preparation and source of screen solutions.

Co-crystallisation of the PX domain at a concentration of 5.1mg.ml^{-1} with the two different phosphatidylinositides was also attempted as described in the methods (5.2.7.4). As only a small amount of each phosphatidylinositide could be purchased only one 96-well screen (SSI and II) could be set up. However, this screen was not successful in producing any crystallisation hits to further direct this work.

In the case of purified Bar domain the Phoenix robot was used to produce 500nl sitting drops. However, despite all 13 screens (Table 5.5) being set up at 3 different protein concentrations (1.6, 5.2 and 10.2mg.ml^{-1}) no crystallisation hit has since been produced to direct further optimisation. It was noted that the majority of drops set up at 10.2mg.ml^{-1} contained heavy precipitation, indicating this concentration to be higher then desired for nucleation.

The 3 MDL 96-HT screens (Table 5.5) were also set up using purified domains that retained their poly-His tags. The tagged domains were purified in approximately the same manner to their cleaved counterparts, the exception being the lack of thrombin cleavage step. The purified His-domains were buffer exchanged into the same pre-crystallisation buffer as their untagged form, with the inclusion of 100mM imidazole to encourage the ordering of the His-tag. Crystallisation trials have been attempted at 5.5mg.ml^{-1} for His-PX and 5.2mg.ml^{-1} for His-Bar, however neither has yet produced a hit.

Screen	Protein ($mg.ml^{-1}$)	Method	Drop size	Condition	Description
SS I	3.4 and 5.1	Sitting	500nl	0.2 M magnesium formate	Spherolites
JBScreen 6	3.4	Hanging	1µl	0.1 M HEPES pH 7.5, 2% PEG 1K, 1.6 M ammonium sulphate	Phase separation
JBScreen 8	2.4 and 5.5	Sitting	500nl	0.1 M Tris-HCl pH 8.5, 2% v/v ios-propanol, 0.1M magnesium sulphate	Phase separation

Table 5.7 Crystallisation conditions producing 'hits' for the apo PX domain.

The two sitting drop hits were produced during robotic screening and the hanging drop hit by manually set up 24-well plate screening.

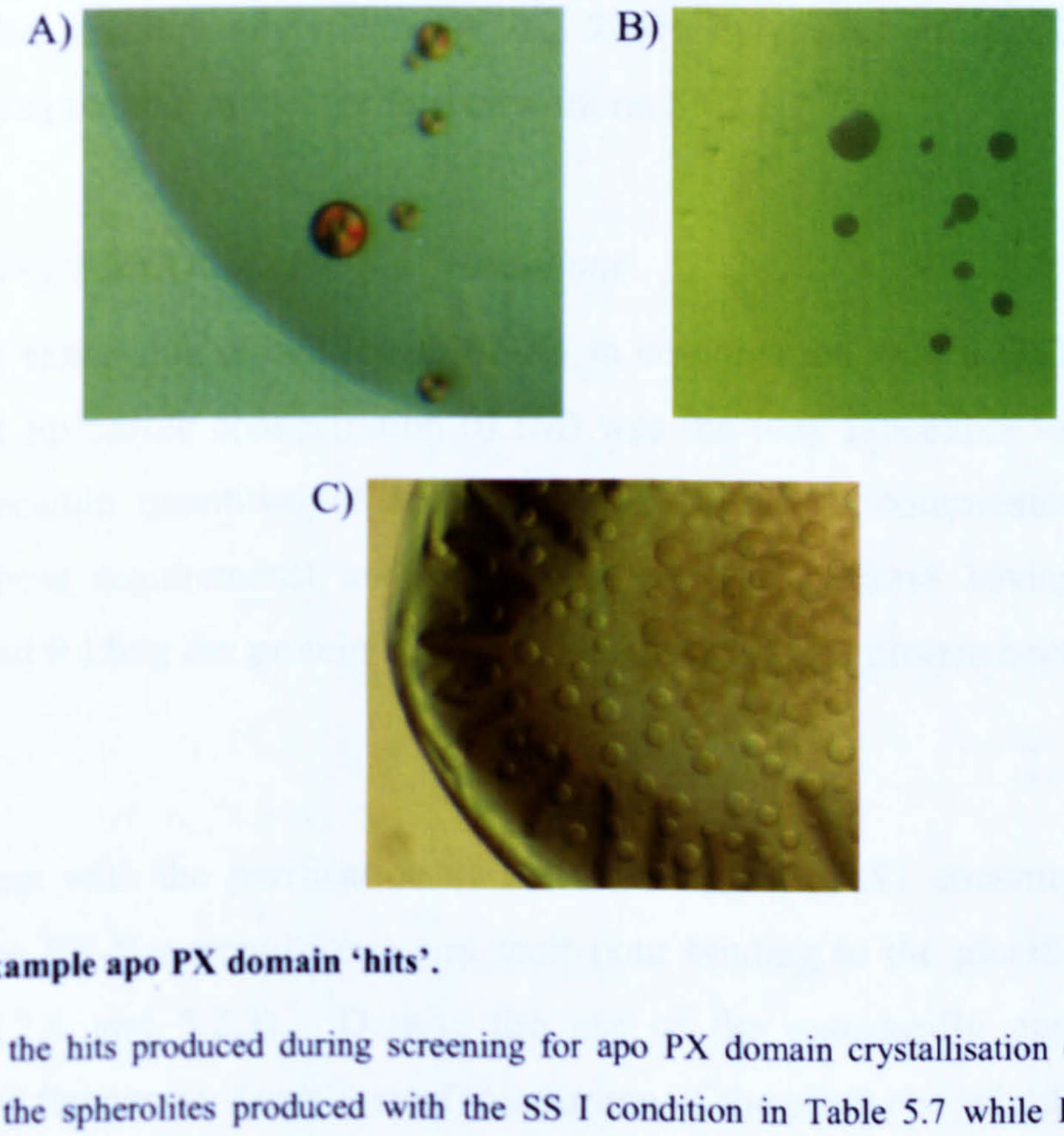


Figure 5.27 Example apo PX domain 'hits'.

Photographs of the hits produced during screening for apo PX domain crystallisation conditions. A) corresponds to the spherolites produced with the SS I condition in Table 5.7 while B) shows these spherolites after Izit treatment. The concentration of the Izit dye with these spherolites has made them darker and is indicative of protein crystals. C) the phase separation produced with the JBScreen 6 condition presented in Table 5.7. Similar phase separation was produced in the other reported condition.

5.4 Discussion

Despite the sorting nexins forming a large, diverse class of proteins involved in various aspects of cellular trafficking (reviewed in Worby and Dixon, 2002) very little structural work on these proteins has previously been reported. At the outset of this study only one crystal structure had been produced, that of yeast SNX3 with and without bound ligand (Zhou *et al.*, 2003). This sorting nexin solely consists of a PX domain. Additionally structures of a small number of further, separate PX domains were available as NMR structures, including SNX1-PX (Zhong *et al.*, 2005). Consequently the scarcity of structural information regarding this important group of proteins led us, in collaboration with Prof. P. Cullen (University of Bristol), to investigate the feasibility of producing milligram quantities of pure, soluble SNX1 protein (or parts there of) with the aim of establishing purification procedures and initiating crystallisation trials. The progress of this work will be discussed with a focus on what can be learnt and applied to any future initiatives. The recently published SNX9 PX-Bar structure (Pylypenko *et al.*, 2007; Pylypenko *et al.*, 2008) is also mentioned as a potential model for further work on SNX1.

5.4.1 Full length SNX1 and PX-Bar truncations

The bacterial expression of full length SNX1 in combination with a GST tag at 16°C and with low imidazole concentration (0.1M) was the only procedure to successfully produce detectable quantities of soluble protein (personal communication Dr. G. Cozier). These requirements are consistent with the bacteria having difficulties expressing and folding the protein and were indicative of this protein being difficult to handle.

A key problem with the purification of the GST tagged SNX1 constructs (both full length and the PX-Bar truncations) was their poor binding to the glutathione column (sections 5.3.2.1 and 5.2.3). Despite the use of the supposedly optimum buffer conditions and temperature only a small proportion of the produced soluble protein was bound and subsequently eluted from the column, suggesting that something was inhibiting the interaction or that the interaction of the protein with the column was intrinsically weak. It seems feasible that the oligomeric state of the protein could

decrease the availability of the GST binding site to the column glutathione. With SNX1 forming dimers and potentially higher order oligomers via the Bar domain, along with the ability of GST itself to dimerise (Dirr, Reinemer and Huber, 1994), there is the potential for large complexes to easily form and these may reduce access to the GST. These effects might be partially abrogated by mixing free glutathione resin with the protein as has been used recently for SNX9 production (Pylypenko *et al.*, 2008). Oligomer formation is consistent with the short retention time seen during subsequent size exclusion chromatography where the protein is eluted in the void volume (Figure 5.15).

It was noted that the expression and purification of a separate Bar domain used a GST-tag to enhance solubility together with a poly-His tag for purification (Weissenhorn, 2005). In this case the expressed protein comprised an N-terminal His-tag directly followed by a GST-tag and finally the domain of interest. This concept was exploited with the GST-PX-Bar-His construct although, due to the construction of the vector, the His-tag was C-terminal and separated from the GST-tag by the Bar protein. Surprisingly, this also resulted in insoluble protein although a similar construct, lacking the His-tag, was fully soluble (sections 5.3.4 and 5.3.5). This again reinforces the instability of this protein even when lacking the large N-terminal region of SNX1. Alternatively, this insolubility could be the result of the C-terminal His-tag disrupting the dimerisation of the Bar domains due to its potential position near this interface. This would also offer an explanation as to why all attempts at refolding failed. As an N-terminal poly-His tag alone also resulted in insoluble protein (Figure 5.18) it may be worth investigating an N-terminal poly-His-GST-tag construct, although this may not avoid issues of higher order oligomer formation. This would require either the modification of the available pGEX vectors or the purchase of other commercial versions and was not feasible during the course of this work.

Upon purification of the GST tagged proteins a second problem developed when removal of the tag was attempted. The presence of a large tag is likely to add structural variability and flexibility with resultant effects on or prevention of crystal packing. Consequently the removal of the tag was deemed to be essential; however upon its cleavage from the full length SNX1 this protein precipitated. The large N-terminal region of SNX1 (residues 1 - 138) contains no predicted secondary structure and may

constitute an intrinsically unfolded region of the protein. Without anything to stabilise this region, such as the GST domain or a physiological partner (such as components of the retromer, Seaman and Williams, 2002) it may form inappropriate interactions leading to the insolubility of the entire protein. Accordingly, it proved very difficult to produce full length SNX1 for crystallisation and a focus on truncated forms including the PX and Bar domains was therefore deemed appropriate.

Attempts to cleave the GST tag from the truncated PX-Bar construct resulted in non-specific cleavage and no useable protein was obtained. The appearance of multiple non-specific cleavage sites within this truncated form may reflect introduced instability and misfolding of the protein. The results produced during the cleavage of the poly-His tag from the PX domain also indicate that there may not be sufficient space between the folded PX domain and the folded GST domain to allow thrombin access. The obstruction of this site would liberate thrombin to attack non-specific sites. It may be desirable to extend the PX domain by a few residues to provide more space between the tag and this domain. This is further discussed in section 5.4.2.2 in relation to the His-PX construct.

Although it has not proved possible to purify either full-length SNX1 or the truncated SNX1 PX-Bar in adequate quantities for crystallisation trials during the course of this thesis, progress has been made that can be used to direct future lines of inquiry. The use of a GST-tag to produce soluble PX-Bar provides a starting point from which the linker between the PX domain and tag can be increased to improve protease accessibility, or the incorporation of an intein mechanism to remove this requirement. Additionally, the inclusion of an N-terminal poly-His tag to improve purification may be worth exploring. The use of other large tags that may enhance solubility, such as Maltose Binding Protein (MBP), may also help avoid the aggregation issue, although may ultimately prove fruitless if the SNX1 proteins itself is poorly folded.

5.4.2 Individual domains

The discrete PX and Bar domains of SNX1 act in a separate, but additive, manner to direct the membrane specificity of the entire protein (Carlton *et al.*, 2004). Consequently, the structural investigation of these regions independently is worth

pursuing as it would provide a lot of information on their separate specificity determinants, leaving only their spatial relationship unclear.

5.4.2.1 BAR domain

The localisation of SNX1 on highly curved membranes and the ability of its overexpression to induce (and stabilise) the tubulation of membranes from the early endosome is dependent on the Bar domain (Carlton *et al.*, 2004). This class of dimerisation domain forms a banana-shaped dimer, which binds to negatively charged membranes via its positively charged concave surface (Gallop and McMahon, 2005). The curvature of this dimer is partly due to the way in which the monomers intersect and has been seen to vary widely across the published Bar structures (Tarricone *et al.*, 2001; Shimada *et al.*, 2007). Knowledge of the structure of the SNX1 Bar domain would show its level of curvature and provide a useful comparison (and model) for other SNX Bar domains, particularly as the SNX1 Bar domain is especially sensitive to membrane curvature (Carlton, 2004). It may also provide further information on the mechanism of membrane tubulation.

The SNX1 Bar domain has been predicted to comprise residues 306 to 496 (Carlton, 2004). It was decided, in discussion with Dr R. Sessions (University of Bristol), to express a construct from residue 296 to 503. This corresponds, at the N-terminus, to the ordered region seen in the dAMP Bar domain structure (Peter *et al.*, 2004) and, at the C-terminus, to a short extension prior to a charged region (Figure 5.7). Previously, most individual Bar domains have been expressed as a GST-tag fusion product (Peter *et al.*, 2004; Weissenhorn, 2005; Masuda *et al.*, 2006), but as there was precedent for soluble Bar expression using a poly-His tag (Casal *et al.*, 2006) this was attempted, and succeeded in producing large quantities of soluble SNX1 Bar domain that could be readily purified.

Disappointingly, despite screening 3 protein concentrations across 528 crystallisation conditions no crystals were obtained. A number of structures from individual Bar domains have been published (Peter *et al.*, 2004; Weissenhorn, 2005; Casal *et al.*, 2006; Masuda *et al.*, 2006) suggesting that despite its elongated, curved nature it is a domain that can be crystallised using standard techniques. Further, the inclusion of a

Factor Xa protease site allowed removal of all extraneous vector encoded amino acids indicating that it is perhaps the Bar domain boundaries used that are preventing crystallisation. Adjustment of these boundaries may be required to produce protein crystals. The soluble expression system and a purification procedure developed during the course of this work will, hopefully, facilitate production of these future variants. It should also be noted that time did not allow a screen of crystallisation temperatures or a complete range of protein concentrations, both of which are variables that could be further explored in future work.

It is possible that this Bar domain was prematurely truncated despite the C-terminal residue (503) extending beyond the predicted domain boundaries although the production of soluble, stable protein suggests the protein is well folded and should be amenable to crystallisation. The size of the SNX1 Bar domain used here matches that seen in the SNX9 structure where this domain continues to the C-terminus (Pylypenko *et al.*, 2007), again supporting the choice of the domain boundaries.

5.4.2.2 PX domain

The PX domain is a small domain known to bind various phosphatidylinositol phosphates (PtdInsP) and to be involved in targeting various proteins to specific membranes (reviewed in Seet and Hong, 2006). The PtdInsP binding specificity varies across the SNX family, with SNX1 binding PtdIns(3)P and PtdIns(3,5)P_2 (Cozier *et al.*, 2002). However, the extra requirement for the Bar domain in membrane binding indicates this as a low affinity PX domain (Zhong *et al.*, 2002; Carlton *et al.*, 2004). Although there is an existing NMR structure of this domain (Zhong *et al.*, 2005), its classification as a low affinity PX domain with dual PtdInsP specificity makes further structural study appropriate.

As previously mentioned, the boundaries of the SNX1 PX domain construct (residues 139 – 269) were based on those successfully shown by Zhong *et al.*, (2005) to be produce soluble and stable protein. This range of residues ensured the entire predicted PX domain was included without the addition of potentially flexible termini which may interfere with crystallisation and also avoided the stretch of charged glutamate residues at the N-terminal (Figure 5.7).

As with the Bar domain construct, the PX domain was designed to include a Factor Xa cleavage site directly at the N-terminus of the domain to allow the removal of the entire tag and hence minimise potential protein flexibility during crystallisation. However, unlike the Bar domain, the addition of Factor Xa resulted in non-specific proteolysis requiring instead the use of the secondary thrombin site to remove the poly-His section of the tag. This resulted in the PX domain having 24 extra N-terminal residues which, while not optimal, was deemed reasonable to move into initial crystal trials. The production of apo PX crystals was targeted with the aim of allowing ligand soaking experiments to take place. These crystallisation trials have so far produced 3 identifiable hits (Table 5.7 and Figure 5.27), one of which was crystalline but failed to diffract. Co-crystallisation of SNX1 PX with PtdIns ligands failed to produce any hits, but due to the poor availability of the commercial di-C_4-PtdIns this can not be considered an in depth screening. A number of high affinity PX domains have been co-crystallised in complex with di-C_4-PtdIns(3)P (Bravo *et al.*, 2001; Zhou *et al.*, 2003), while this ligand was successfully soaked into the SNX9 PX-Bar crystals (a low affinity PX domain, Pylypenko *et al.*, 2007). The lack of co-crystallisation hits may be due to the intrinsic low affinity of the SNX1 PX domain for its PtdIns ligand but further screening is required. If no further source of the di-C_4-PtdIns ligands can be found it may be necessary to use versions with longer "tails" such as C_8 which is still water soluble or the phosphoinositide head group itself, both commercially available from Echelon (Utah, USA).

The small number of identified crystallisation conditions may also be due to the construct used. The excess amino acids retained at the N-terminus due to the use of the thrombin cleavage site may have added a region high in flexibility which makes crystal packing more difficult. The apparent inability of the Factor Xa protease to access its cleavage site may be indicative of this site being shielded by the structure. The NMR SNX1-PX structure indicates that the constructs N-terminus is flexible, but its position at the end of a β-strand and potential to pack against other secondary structures may result in poor accessibility for the proteases. This would be particularly true when the GST domain is attached as in the previous, larger constructs. As Zhong *et al.*, (2005) used self-cleavage by an incorporated intein to remove the tag from their SNX1 PX construct they did not require access by a secondary factor, negating this structural

factor. The structure of SNX9 is 34 residues longer at the N-terminus then the PX boundary used in these studies (Pylypenko *et al.*, 2007). Although these extra residues largely fold back on to the Bar domain it may be possibly to significantly extend the N-terminus and still retain a compact domain. This would be worth investigating, with an N-terminal extension to the region of residue 121 being a potential starting point, as this would avoid the string of charged residues and the inclusion of prolines near the beginning (start at ATN, Figure 5.7). This would also be an appropriate extension to trial for the expression of the entire PX-Bar region, particularly as these N-terminal residues may be involved in interactions with the Bar domain, as seen for SNX9 (Pylypenko *et al.*, 2007). The expression and purification procedures established in this thesis will hopefully assist in this process once the constructs have been made. It would also be potentially useful to determine the domain boundaries (of both the PX and Bar domains) via limited proteolysis, this however would require improved production of the full length SNX1 protein.

Additionally, the NMR structure shows a large amount of flexibility in a loop region that forms part of the PtdInsP binding pocket (Zhong *et al.*, 2005). This flexibility may make the production of an apo crystal form significantly more difficult then if it was stabilised. As this region is expected to become more ordered when bound to its ligand (Zhong *et al.*, 2005) it might be appropriate to focus future work on co-crystallisation with PtdIns(3)P and PtdIns(3,5)P_2, although the success of soaking ligand as seen in the SNX9 structure (Pylypenko *et al.*, 2007) indicates that the production of apo crystals is still a worthwhile aim.

While the recent structure of SNX9 PX-Bar has provided fresh insight into the mechanism by which membrane specificity is achieved in this sorting nexin, there are significant differences between the function and specificity of SNX9 compared with SNX1. Consequently, the importance of structural studies on SNX1 has not been diminished but may, as described above, be advanced by the knowledge gained from the successful SNX9 structure.

Chapter 6 Investigation of Binding Partners for TFL1 and FT

6.1 Introduction

The *Arabidopsis* proteins produced by the genes *TERMINAL FLOWER 1* (TFL1) and *FLOWERING LOCUS T* (FT) act antagonistically to control the timing of floral meristem formation. These closely related proteins are members of the phosphatidylethanolamine-binding protein (PEBP) family, the most highly studied of which is the Raf kinase inhibitor protein (RKIP). The crystal structures of both TFL1 and FT have been produced previously (Ahn *et al.*, 2006), and in this study phospho-amino acid soaking and co-crystallisation experiments have been undertaken as an initial extension of these structural studies towards complexes with their binding partners.

6.1.1 RKIP/PEBP family

The RKIP/PEBP family is a widespread, highly conserved protein family found in many organisms, including humans (Tohdoh *et al.*, 1995), flowering plants (Bradley *et al.*, 1997), *Drosophila* (Pikielny *et al.*, 1994) and *E. coli* (Serre *et al.*, 2001). They were initially named as phosphatidylethanolamine binding proteins (PEBP), due to an *in vitro* affinity for PE (Bernier, Tresca and Jolles, 1986). Now, they are more commonly referred to as Raf kinase inhibitor proteins (RKIP, Yeung *et al.*, 1999), due to the action of a rat form as described shortly. In general however, the mode of action of this ubiquitous family is unclear.

A number of RKIP/PEBP family members have had their structures solved, including 3 mammalian forms (human, bovine and mouse: Banfield *et al.*, 1998; Serre *et al.*, 1998; Simister, Banfield and Brady, 2002), 3 plant forms (*Antirrhinum* CEN, *Arabidopsis* TFL1 and FT, Banfield and Brady, 2000; Miller, 2004; Ahn *et al.*, 2006) and two *E. coli* members (Serre *et al.*, 2001). All structures show a high level of similarity with a large central β-sheet, an anion binding site at one end and an external loop likely to be involved in determining some of the specificity for each protein (Figure 6.1).

The anion binding site is created from a number of highly conserved residues which together form a surface exposed pocket (Figure 6.5). A number of crystal structures of PEBP/RKIP members contain phosphorylated ligands (ie PE), the phosphate analogue cacodylate or an acetate ion bound in this pocket (Banfield *et al.*, 1998; Serre *et al.*, 1998), indicating a potential role in binding phosphorylated ligands. Moreover, mutations within this binding pocket have been shown to have a significant negative effect on the ability of RKIP to interact with the Raf-1 kinase, an interaction normally facilitated by the phosphorylation of Raf-1 (Rath *et al.*, 2008). The (non peer reviewed) crystal structure of human RKIP containing a phospho-tyrosine ligand bound in this pocket (Simister, 2004), strongly supports the hypothesis that this pocket represents a phospho-amino acid binding site.

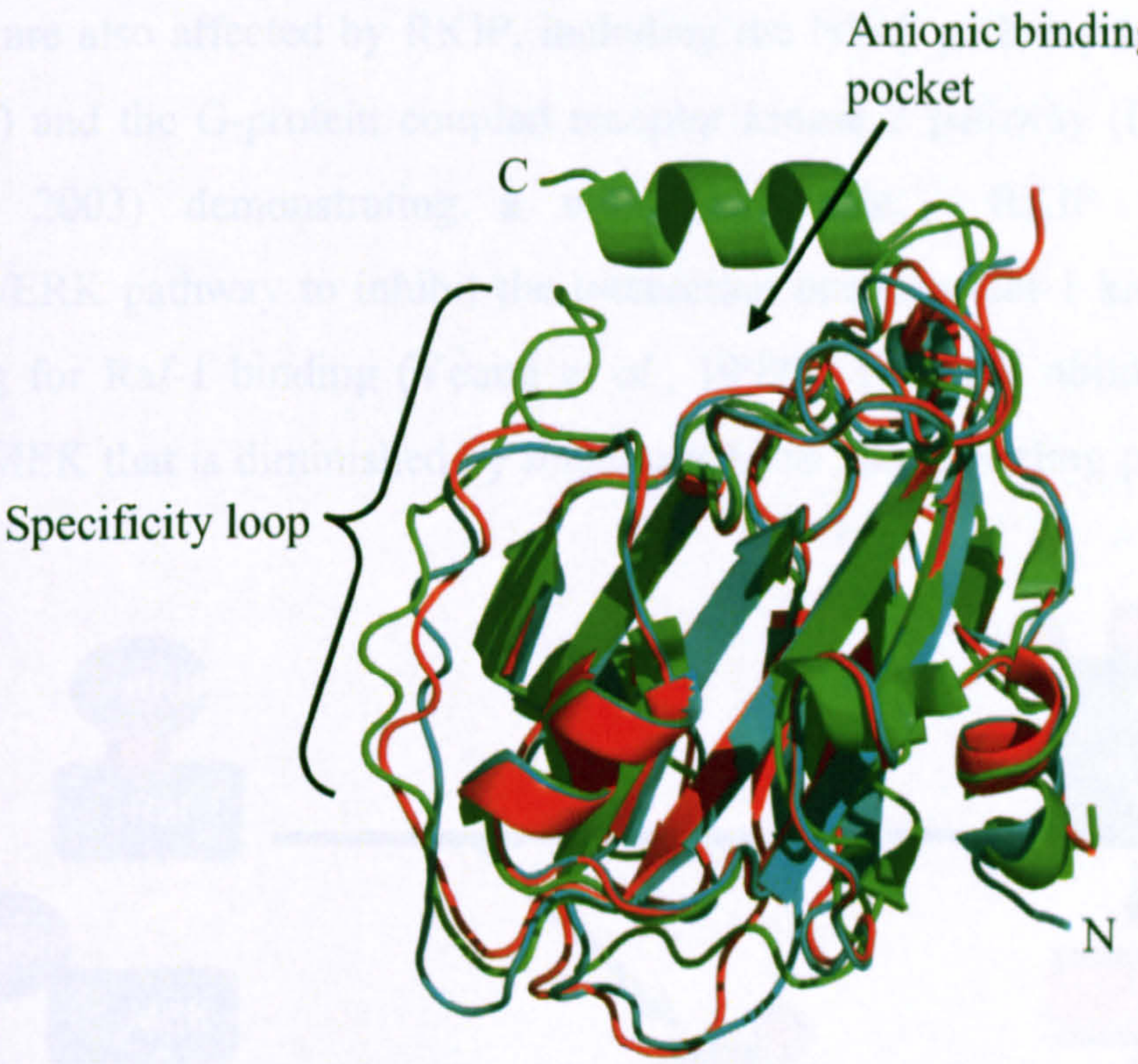


Figure 6.1 Overlay of TFL1, FT and human RKIP with salient features indicated.

Cartoon representation of TFL1 (cyan), FT (red) and human RKIP (green). PDB access codes are 1WKO, 1WKP and 2QYQ respectively.

6.1.2 Role in kinase signalling pathways

Evidence is accumulating that indicates a general role for the RKIP/PEBP family in the regulation of various signalling pathways. The correct growth and development of multicellular organisms is often dependent on the appropriate response to an extracellular stimulus such as growth factors, hormones and stress. This response often involves the induction or inhibition of a particular transcription event, requiring an extracellular stimulus to be converted into an event at the nucleus. This is often facilitated by the sequential activation of one kinase by another (a kinase cascade), culminating with the activation of a transcription factor or regulator.

Many studies now indicate the importance of some mammalian forms of the RKIP family in inhibiting the Raf/MEK/ERK mitogen activated protein kinase (MAPK) pathway (Yeung *et al.*, 1999; Yeung *et al.*, 2000). This pathway is involved in controlling cell growth in response to hormones and growth factors. Additional kinase pathways are also affected by RKIP, including the NFκβ pathway (Chen, Demers and Shi, 2002) and the G-protein coupled receptor kinase 2 pathway (Lorenz, Lohse and Quitterer, 2003) demonstrating a widespread role. RKIP functions in the Raf/MEK/ERK pathway to inhibit the interaction between Raf-1 kinase and MEK by competing for Raf-1 binding (Yeung *et al.*, 1999). It is this ability of RKIP to out compete MEK that is diminished by mutations to its anion binding pocket (Rath *et al.*, 2008).

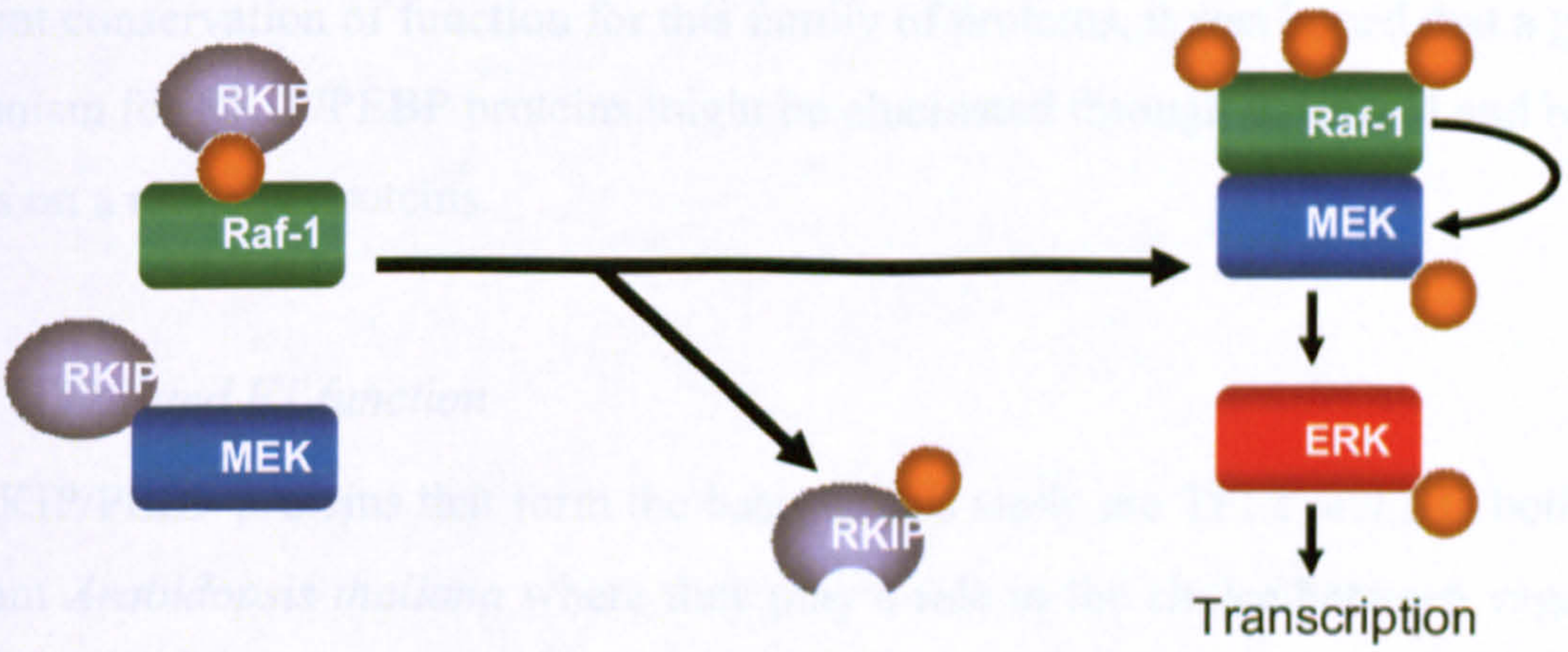


Figure 6.2 Simplified RKIP mechanism.

When RKIP is dissociated, Raf-1 and MEK can interact, MEK is phosphorylated and activated by Raf-1, MEK in turn phosphorylates and activates the transcription factor ERK. RKIP/ERK interaction not shown for clarity. Phosphorylation shown as orange balls.

Binding to partially phosphorylated Raf-1 kinase allows RKIP to act as a buffer, soaking up Raf-1 until full activation is achieved. The level of RKIP can control the threshold level at which the Raf/MEK/ERK pathway activates. This inhibition can also be relieved by phosphorylation of RKIP itself (Corbit *et al.*, 2003). However molecular level details of this mechanism are not fully understood, partially due to the difficulty in obtaining the interacting proteins for structural studies.

Additionally, work on a tomato member of RKIP/PEBP family, SELF-PRUNING (SP), also supports a role for SP within a signalling pathway involved in determining the growth pattern of tomato plants (Pnueli *et al.*, 2001). SP (a close homologue to TFL1 and FT) was shown to interact with a serine/threonine kinase (SPAK), a bZIP (basic region/leucine zipper) transcription factor (SPGB) and a 14-3-3 isoform (Pnueli *et al.*, 2001). The interaction of SP and SPAK was shown to require that SPAK be phosphorylated, just as RKIP requires that Raf-1 be phosphorylated for complex formation. Moreover, bZIP transcription factors are known to require phosphorylation to be transcriptionally activated (Jakoby *et al.*, 2002) and 14-3-3 proteins can act as scaffolds for the assembly of signalling complexes (Roberts, 2003; Ferl, 2004). The *Arabidopsis* RKIP/PEBP proteins, TFL1 and FT, can also replace SP in many of these interactions *in vitro* (Pnueli *et al.*, 2001). While not all of these interactions may be physiologically relevant, in general these observations are consistent with a primary role for RKIP/PEBP proteins as components of signalling pathways. Because of this apparent conservation of function for this family of proteins, it was hoped that a general mechanism for RKIP/PEBP proteins might be elucidated through structural and binding studies on a range of proteins.

6.1.3 TFL1 and FT function

The RKIP/PEBP proteins that form the basis of this study are TFL1 and FT, both from the plant *Arabidopsis thaliana* where they play a role in the choice between vegetative growth and floral development (reproductive growth). The normal pattern of growth in *Arabidopsis* is classified as indeterminate, where the apical meristem grows indefinitely and produces floral meristems from the periphery (Figure 6.3). The choice between vegetative and reproductive growth is dependent on various environmental signals sensed and relayed into the cell by a large number of genes. Under appropriate

conditions these result in the formation of an inflorescence, the reproductive structure. Mutagenesis and over-expression studies suggest these two highly homologous proteins (56% sequence identity, r.m.s.d of 0.77Å over 157 Cα atoms) act in a generally antagonistic function to control the timing of inflorescence formation. The flowering of *ft/ft* mutants is delayed indicating FT as a floral promoter (Kobayashi *et al.*, 1999), while *tfl1/tfl1* mutants show early flowering and reduced vegetative growth, indicative of TFL1 as an inhibitor of inflorescence formation (Bradley *et al.*, 1997, Figure 6.3). The respective overexpression studies show a delay in flowering when TFL1 is overexpressed (Bradley *et al.*, 1997) and an enhancement of flowering when FT is overexpressed (Kardailsky *et al.*, 1999; Kobayashi *et al.*, 1999), indicative of their antagonistic effect. While slight variations in plant phenotypes produced by loss of function and over-expression mutants do indicate some independent activities, separate FT- and TFL1-specific binding partners have yet to be identified.

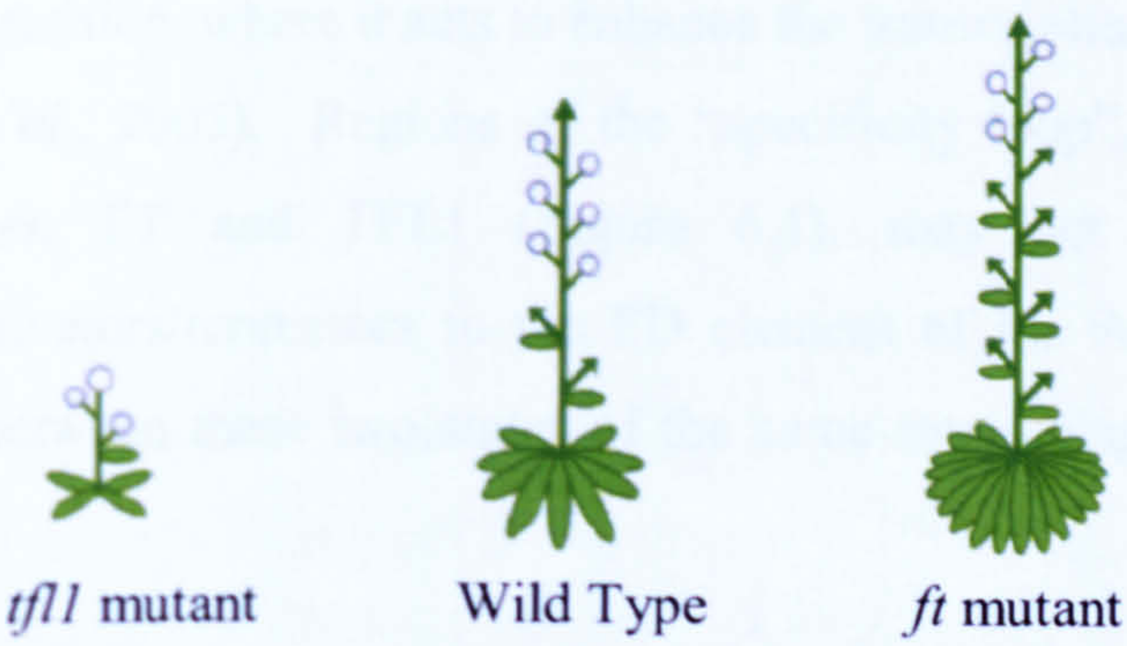


Figure 6.3 Phenotypes of TFL1 and FT mutants.

The *tfl/tfl1* mutation results in much earlier flowering, while the *ft/ft* mutant allows TFL1 to dominate, resulting in far later flowering. Leaves = ellipses, stems/shoots = lines, flowers = circles, and shoot meristems = arrows. Adapted from Hanzawa, Money and Bradley, 2005.

TFL1 and FT appear to represent a convergence point for the various signals that control the timing of the vegetative to reproductive switch. The vernalisation (low temperature) pathway controls the level of the floral inhibitor FLOWERING LOCUS C (FLC) that in turn prevents the expression of pro-reproductive genes, including *FT* (Henderson and Dean, 2004, Figure 6.4). The photoperiod (day-length) regulates the activity of CONSTANS (CO) that, during long days accumulates and activates the expression of pro-reproductive genes, with a major effect being the upregulation of *FT*

(Samach *et al.*, 2000, Figure 6.4). TFL1 negatively regulates the floral activator genes *APETALA 1* (AP1) and LEAFY (LFY) (Liljegren *et al.*, 1999) while FT activates them (Samach *et al.*, 2000), meaning the balance between these two proteins has a large effect on flowering time (Figure 6.4). However, unlike many of the other genes regulated by FLC and CO, FT and TFL1 are not transcription factors so the mechanism by which they perform their regulatory functions has been unclear until recently. The identification of the bZIP transcription factor FD as an interacting partner of both TFL1 and FT (Abe *et al.*, 2005; Wigge *et al.*, 2005) goes some way to explaining the antagonistic relationship of these proteins. The FT:FD interaction was also shown to be dependent on the retention of a proposed Ca^{2+}-dependent protein kinase (CDPK) phosphorylation site (either a threonine or serine) at the C-terminus of FD (Abe *et al.*, 2005). This is consistent with the observed dependence of RKIP (Rath *et al.*, 2008) and SP (Pnueli *et al.*, 2001) on phosphorylation for binding to their respective kinases. Immunoprecipitation studies also showed FT is recruited to the AP1 promoter region in an FD-dependent fashion, where it acts to enhance the transcriptional activation activity of FD (Wigge *et al.*, 2005). Regions of the "specificity loop", the major region of difference between FT and TFL1 (Figure 6.1), may act to recruit different transcriptional activators/repressors to the FD element of the AP1 promoter or may induce a switch between these two states of the same interacting partner (Ahn *et al.*, 2006).

The expression of FT as controlled by CO is restricted to the phloem tissue of leaves (Takada and Goto, 2003), while FD appears to be constitutively expressed in the shoot apex (Abe *et al.*, 2005; Wigge *et al.*, 2005). Consequently, when FT is expressed in response to the increase in day length it must relocate to the shoot apex where FD defines the location of its effect. This mobile signal has recently been discovered to be the FT protein itself rather then mRNA as previously thought (Corbesier *et al.*, 2007; Jaeger and Wigge, 2007; Mathieu *et al.*, 2007). This long range protein movement has also be demonstrated for the rice FT homolog (Tamaki *et al.*, 2007) and for TFL1, which is also expressed at a spatially separate site from its mode of action (Conti and Bradley, 2007).

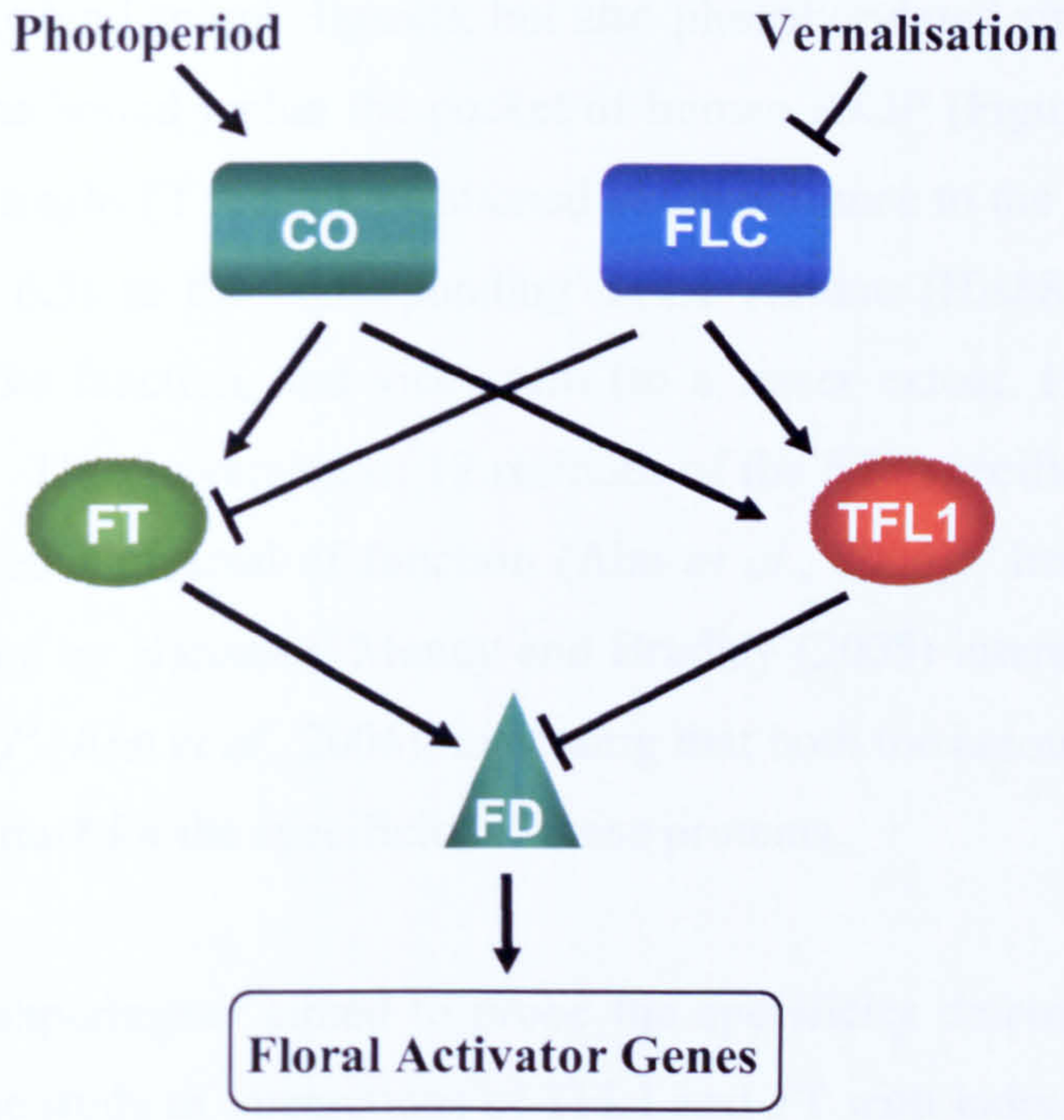


Figure 6.4 Simplified mechanism for the control of flower timing in *Arabidopsis*.

Floral activator genes include AP1 and LFY, with arrows indicating activation/induction and blunted lines indicating repression.

Close homologues of FT and TFL1 with similar roles have been found across the flowering plant family including separate homologues in rice (Kojima *et al.*, 2002; Nakagawa, Shimamoto and Kyozuka, 2002), and TFL1 homologues in the snapdragon (CEN, Amaya, Ratcliffe and Bradley, 1999), tomato (SP, Pnueli *et al.*, 1998) and pea (DET and LF, Foucher *et al.*, 2003). The apparent conservation and importance of this growth determination mechanism across flowering plants makes the study of this pathway of great interest.

The proposed anion binding pockets of FT and TFL1 (and the *Antirrhinum* homolog CEN, Banfield and Brady, 2000) are more accessible then their mammalian counterparts, due to the absence of a C-terminal helix that is present in RKIP (Ahn *et al.*, 2006, Figure 6.1). Despite being crystallised in the presence of phosphate analogues, in the respective crystal structures these regions are observed to only bind solvent molecules (Banfield and Brady, 2000; Ahn *et al.*, 2006). Nevertheless, the residues within this pocket are consistent with an anion binding function (Figure 6.5), with the FT pocket being slightly enlarged compared with TFL1 although not significantly enough to easily explain their alternative functions. These pockets could

not only feasibly bind anionic ligands, but also phosphorylated amino acids such as the phospho-tyrosine bound within the pocket of human RKIP (Figure 6.5). Interestingly the switch of a single FT residue positioned at the entrance to the anion binding pocket (Tyr85, Figure 6.5) to the corresponding TFL1 residue (His88) results in FT with largely TFL1-like function, and vice versa (to a lesser extent, Hanzawa, Money and Bradley, 2005). The conversion of 12 residues of the FT "specificity loop" to those of TFL1 also caused a reversal of function (Ahn *et al.*, 2006). Importantly the residue (His88) identified by Hanzawa, Money and Bradley (2005) interacts with those of the "specificity loop" (Ahn *et al.*, 2006), indicating that both the anion pocket and this loop region are important for the specificity of these proteins.

The following experiments aimed to probe the specificity determinants in TFL1 and FT. Through the study of interactions of TFL1 and FT with individual phospho-amino acids, particularly at a structural level, it was hoped to indirectly determine some of the features present in their respective protein ligands. This might generally be used to further elucidate the role of FT and TFL1 in the timing of flowering in *Arabidopsis* and hence unravel the general role played by the proteins from the RKIP/PEBP family.

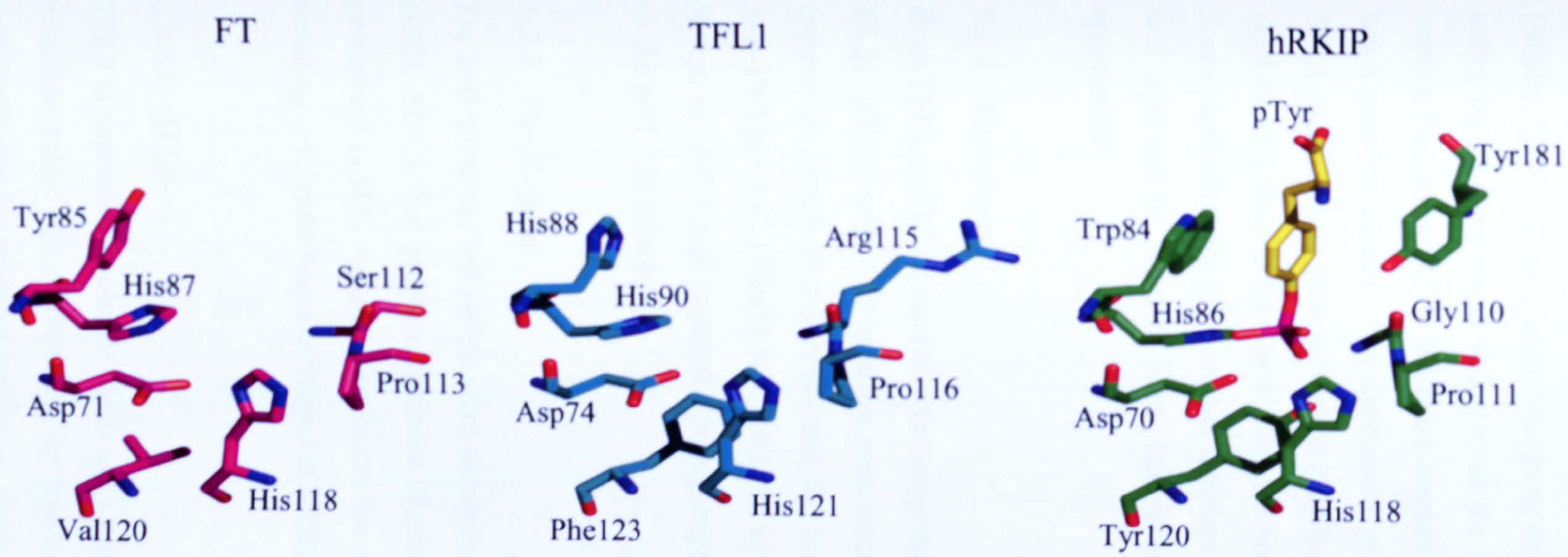


Figure 6.5 Proposed phospho-amino acid binding pockets of FT and TFL1 in comparison with the hRKIP/pTyr complex.

Selected amino acids forming the binding pockets are shown in the same orientations, coloured in accordance with the overlay in Figure 6.1, with the addition of phospho-tyrosine shown in yellow. The pockets are aligned to be open to solvent at the top. FT (1WKP) and TFL1 (1WKO) are published structures (Ahe *et al.*, 2006) and the hRKIP/phospho-tyrosine structure is unpublished (Simister, 2004, 2QYQ). Adapted from Miller, 2004.

6.2 Methods

Both TFL1 and FT had been previously cloned into the expression vector pET-28a (Novagen), which attaches a thrombin cleavable hexa-His tag to the N-terminus (Miller, 2004). The FT used throughout this study has had all its cysteines (3) removed either by mutagenesis to serine or as the result of a small C-terminal deletion. This version has proven to be the most amenable to study (Miller, 2004). The TFL1 used has a native sequence. The expression, purification and initial crystallisation of both TFL1 and FT have been published elsewhere (Miller, 2004 and Ahn *et al.*, 2006) and shall be described in brief here. The methods used in this chapter are presented below. Some methods from the previous chapters (Chapter 2 and Chapter 5) were also used and will not be repeated here, but referenced by their section numbering where appropriate.

6.2.1 Expression

Both TFL1 and FT constructs were supplied as -80°C stocks of pre-transformed *E. coli* BL21 (DE3) cells. Expression was undertaken as described for GRHPR, except expression lasted for 5-6 hours and the resulting cell pellets were washed in 20mM Tris-HCl pH 7.4, 150mM NaCl, 0.02% w/v sodium azide (NaN_3). Cell pellets were frozen at -20°C until required. During expression kanamycin (50$\mu g.ml^{-1}$) was present as the pET-28a selector.

6.2.2 Purification

For both proteins, cell pellets were resuspended in 20mM BisTris-HCl pH 6.5, 100mM NaCl, EDTA-free protease inhibitor (Roche) and sonicated on ice. Sonication was performed at 40% Ampl, 10sec bursts for a total of 200sec (VibraCell, Sonics). The resulting cell lysate was centrifuged at 25000g, 4°C for 30min.

6.2.2.1 TFL1

TFL1 is expressed insolubly and is only present in the pellet after centrifugation. The insoluble pellet was resuspended in 20mM BisTris-HCl pH 6.5, 100mM NaCl, 4M urea, followed by recentrifugation at 25000g, 15°C for 30min. This wash step was

executed twice. The resulting pellet (still containing TFL1) was resuspended in Denaturing PEBP Ni^{2+} Loading Buffer (50mM BisTris-HCl pH 6.5, 250mM NaCl, 50mM imidazole, 20mM ethanolamine, 8M urea, 10% (v/v) glycerol and 0.02% (w/v) NaN_3). A sonicating water bath was used to assist the resuspension of denatured protein. The denatured protein was clarified by centrifugation as above. The solubilised, denatured protein was loaded on to a 5ml Ni^{2+} affinity column and washed until a flat baseline was achieved. Bound protein was then eluted using a gradient to 1M imidazole. The fractions containing denatured TFL1 were identified, and refolding was accomplished by a 50-fold rapid dilution into Native PEBP Ni^{2+} Loading Buffer (50mM BisTris-HCl pH 6.5, 250mM NaCl, 50mM imidazole, 10% (v/v) glycerol and 0.02% (w/v) NaN_3). This was then spun at 25000g, 4°C for 30min and the supernatant loaded onto a new non-denatured Ni^{2+} affinity column, equilibrated in the same buffer. Elution of bound protein was again achieved with an imidazole gradient to 1M. Fractions predominantly containing TFL1 were pooled and dialysed (14kDa cut-off dialysis tubing) at 4°C, using two changes of a 100x volume of MonoS Loading Buffer (50mM BisTris-HCl pH 6.5, 20mM NaCl and 0.02% NaN_3). The protein concentration was determined by A_{280} (MW = 20158Da, $\varepsilon = 9530 M^{-1}.cm^{-1}$) and the His-tag was removed by the addition of 10U thrombin per mg of protein with incubation at 18°C for 16hrs. Cleaved TFL1 was purified away from thrombin and uncleaved protein by ion exchange chromatography using a cation exchange resin (MonoS, Amersham Biosciences). A gradient of 20mM - 500mM NaCl was used to elute protein from the MonoS column. Pure, cleaved TFL1 was then used for crystallisation.

6.2.2.2 FT

A fraction of the expressed FT is soluble and hence the clarified supernatant was loaded directly onto a native Ni^{2+} affinity column. Bound protein was eluted with a 0 - 1M imidazole gradient, pooled and protein concentration determined by A_{280} (MW = 19,809Da, $\varepsilon = 20340 M^{-1}.cm^{-1}$). Cleavage of the His-tag was accomplished by the addition of thrombin ($10U.mg^{-1}$) and incubation for 16hrs at 18°C. Separation of cleaved FT from uncleaved protein and thrombin was performed by gel filtration (Superdex 75 16/60) in FT Stabilisation Buffer (20mM CHES pH 9.5, 100mM NaCl). Fractions containing pure FT were pooled and this protein used for further experimentation.

6.2.3 Crystallisation

Both TFL1 and FT have previously been crystallised (Miller, 2004). The hanging drop vapour diffusion method was used to set up crystal trials around those conditions previously shown to be successful.

6.2.3.1 TFL1 crystallisation

TFL1 was buffer exchanged into 20mM BisTris-HCl pH 6.5 and concentrated to either 10mg.ml^{-1} or 15mg.ml^{-1} using a 5kDa cut-off Centricon (Amicon). The published crystallisation conditions of TFL1 were; 10mg.ml^{-1} in 0.3M $(NH_4)_2SO_4$, 22% w/v PEG 5K MME and 0.1M Cacodylate pH 5.0 (Ahn *et al.*, 2006). In addition, quality crystals had been produced with 0.1M MES pH 6.5, 0.2M $(NH_4)_2SO_4$ and 30% w/v PEG 5K MME (personal communication, Dr D. Miller, University of Bristol). Hanging drop vapour diffusion crystallisation trays were set up as described in section 2.3.15.2. A range of conditions, given in Table 6.1 were used in order to maximise the chances of reproducing the published crystals

Crystallisation conditions	
0.1M Cacodylate pH 5.0	0.1M MES pH 6.5
18% - 30% PEG 5000 MME	22% - 32% PEG 5000 MME
0.2M or 0.3M $(NH_4)_2SO_4$	0.2M or 0.3M $(NH_4)_2SO_4$
+/- 10mM ammonium phosphate	+/- 10mM ammonium phosphate

Table 6.1 Screening conditions used for TFL1 crystallisation.

Based on successful conditions determined in Miller, 2004.

6.2.3.2 FT crystallisation

FT was retained in the Stabilisation Buffer and concentrated to 10mg.ml^{-1} using a 10kDa cut-off Centricon. The original FT crystallisation was undertaken at 10mg.ml^{-1} in 0.14M $(NH_4)_2SO_4$, 38% w/v PEG 5K MME and 0.1M MES pH 7.1 (Ahn *et al.*, 2006) but was poorly reproducible (personal communication Dr D. Miller). Attempts were made to reproduce and improve on these crystals by re-optimising around this condition (Table 6.2).

Buffers	0.1M MES pH 7.0, 7.1, 7.2 0.1M HEPES pH 7.0, 7.1, 7.2
Precipitant	30% - 40% w/v PEG 5K MME
Salts (0.1 - 0.2M)	$(NH_4)_2SO_4$ $LiSO_4$ Mg_2SO_4

Table 6.2 Initial screening conditions for FT crystallisation.

Based on successful conditions determined in Miller, 2004.

Additionally, a previously unoptimised crystal screen hit (0.1M sodium acetate pH 4.6, 14% PEG 4K) was chosen for further screening. It was hoped that the significantly different conditions might produce a more stable and reproducible crystal for soaking experiments. The additive screen (Hampton Research) was used to attempt to improve any conditions were crystals formed (section 2.3.15.2)

6.2.4 Phospho-amino acid experiments

6.2.4.1 Soaking

Visually good quality apo crystals were selected for soaking experiments. As it was not clear which phospho-amino acid may bind, all 3 (phospho-serine (p-Ser), -tyrosine (p-Tyr), and –threonine (p-Thr)) were attempted. Stock solutions (1M) of each phospho-amino acid were made in a buffer as close to that of the protein being investigated as possible. Small volumes of mother liquor were made, matching each crystal reservoir condition except for the addition of 0.1M phospho-amino acid. High concentrations of phospho-amino acid were used as the interaction with the protein was expected to be weak. Soaking experiments were implemented in 2 ways; normal and gentle. Normal soaking involved the direct transfer of a single crystal to a 2µl phospho-amino acid/mother liquor drop on a fresh cover slip. Gentle soaking was used to try to reduce the potential "shock" of transfer. Firstly the selected crystal was transferred to a fresh 2µl drop of its original mother liquor. Next 1µl of phospho-amino acid/mother liquor was added, followed by the removal of an equivalent volume from the drop. This was then left for 1hr at 18°C and the process repeated (more than 4

times), after which it was assumed the crystal was now fully in soak solution. In both cases the soaking drop was resealed over mother liquor to prevent the drop drying out. Crystals remained in the soak solution for between 24hrs and 2 months, after which they were either plunge frozen in liquid nitrogen for storage or directly tested for X-ray diffraction.

6.2.4.2 Co-crystallisation

Co-crystallisation of FT with the 3 phospho-amino acids was also attempted. Protein was concentrated as before to 10mg.ml^{-1} and the phospho-amino acid added to various concentrations (15mM, 30mM and 50mM). The solution was stored on ice for 30min and centrifuged (18,000g, 5 min) before crystal tray set up. Conditions successful for the apo-crystal were trialled, as was the Structure Screen I and II HT-96 (MDL). JBScreen 3 was attempted with the hanging drop method after a hit relating to this screen was identified in the MDL screen. Further improvements of potential crystals have been attempted with the additive screen (Hampton Research) and by lowering the crystallisation temperature to 4°C. All were set up by hand as 1µl mother liquor with 1µl protein drops.

6.2.5 Data collection, processing and solving

Data was collected at the Daresbury SRS, Warrington as described in section 2.3.16. Between 5%-20% glycerol was used as a cryoprotectant depending on the well conditions of the crystal. These cryoprotectant solutions also contained an equivalent concentration of the phospho-amino acid under investigation.

Data processing, solving and refinement were performed largely as described in sections 2.3.17 to 2.3.20. A monomer of each protein was extracted from their PDB entry (TFL1 - 1WKO and FT - 1WKP) and used for molecular replacement. Structures were refined to a point where the presence or absence of something in the binding pocket could be determined. TLS was not used during the refinement of these structures.

6.2.6 FD

6.2.6.1 Expression

The cloned, full length FD in the expression vector pET-16b (Novagen) was supplied by Dr T. Joseph-Horne (University of Bristol) and expression of FD from this vector results in the attachment of 38 additional (thrombin cleavable) amino acids to the N-terminus, including a deca-His tag. Expression was done as for TFL1 and produced insoluble protein.

6.2.6.2 Purification

The cell pellet of FD expression was resuspended in FD Resuspension Buffer (50mM HEPES pH 7.2, 200mM NaCl, 20mM imidazole, 10% v/v glycerol) with EDTA-free protease inhibitor, sonicated and centrifuged as standard. The resulting pellet was resuspended in the same buffer and respun. This washed pellet was finally suspended in FD Denaturing Ni^{2+} Loading Buffer (50mM HEPES pH 7.2, 200mM NaCl, 8M urea) and centrifuged again. The denatured, solubilised protein was loaded on to a 5ml Ni^{2+} affinity column, washed and bound protein was eluted using an imidazole gradient. Fractions containing FD were identified by SDS-PAGE and used in refolding experiments

6.2.6.3 Refolding attempts

Tests to confirm refolding included concentrating the sample or buffer exchanging into Ni^{2+} column buffer. In both case incorrectly or partially folded protein would be expected to aggregate and precipitate.

Dilution and dialysis

Refolding was initially attempted using both a 50-fold rapid dilution into 3 different 'simple' buffers, 50mM Tris-HCl pH 8.0/HEPES pH 7.2/BisTris-HCl pH 6.5, 200mM NaCl and 10% glycerol, or dialysis into the same buffers. Dialysis was performed both at room temperature and 4°C, using dialysis buttons as described in section 5.2.6.1. In the case of rapid dilution, successful refolding was monitored by Trichloroacetic acid (TCA) precipitation and SDS-PAGE. Briefly, a sample of the refold dilution was

mixed with 1/5th the volume of 3.2M TCA and incubated for 15min. Centrifugation at 18,000g for 10min was followed by washing the pellet in ethanol and re-centrifugation. The resulting pellet was resuspended in SDS sample loading buffer, boiled and run on SDS-PAGE.

Refolding screen

The commercially available protein refolding screen (MDL) was undertaken as described in section 5.2.6.2.

On-column refolding

On-column refolding was attempted following the method of Oganesyan, Kim and Kim, (2005). Briefly, denatured protein was loaded onto a Ni^{2+} affinity column as for normal denaturing purification. This was then washed, using a shallow gradient, into buffer containing 0.1% Triton X-100 and no urea. This detergent wash is designed to prevent misfolding upon removal of the denaturant. The Triton was removed by quickly washing the column with buffer containing 5mM β-cyclodextrin to remove the detergent and promote correct folding. Finally the column was washed into 100mM HEPES pH 7.2, 100mM NaCl, 1mM β-mercaptoethanol. 'Refolded' protein could then be eluted with a gradient to 1M imidazole.

6.3 Results

6.3.1 Overview

The following describes the attempts to produce crystal structures of TFL1 and FT in complex with their potential partners. Apo crystal soaking and co-crystallisation experiments with phospho-amino acids are presented. In total 6 data sets were collected and solved, none of which contained ligand. The purification and refolding efforts carried out on FD, FT's proposed protein interacting partner, are also presented. This is followed by a discussion on potential reasons why no complexes could be produced and proposals for future work.

6.3.2 Purification

6.3.2.1 TFL1

The expression of TFL1 from the pET-28a vector results in the production of insoluble protein, despite past efforts to alter this (Miller, 2004). However, a reasonably simple refolding and purification procedure existed and was used in this thesis.

Initial purification of TFL1 exploits the fact that poly-His tags do not rely on a specific 3-dimensional structure for their affinity, and can therefore be successfully used to purify denatured protein. Denaturing the insoluble TFL1 with 8M urea resulted in a supernatant that could be loaded onto a Ni^{2+} affinity column equilibrated in denaturing buffers. Bound protein eluted with approximately 200mM imidazole, resulting in a significant A_{280} peak that corresponded to reasonably pure TFL1. Fractions that were concurrent with this were pooled and refolded. Refolding was carried out by rapid dilution into non-denaturing Ni^{2+} column buffer. After centrifuging this refolded protein, a small pellet was visible, indicating that refolding had not been 100% successful. Reloading the resulting supernatant onto a non-denaturing Ni^{2+} affinity column and its subsequent elution both concentrated and further purified the refolded protein (Figure 6.6 and Figure 6.7a). Refolded TFL1 was eluted by approximately 0.4M imidazole which had to be removed by dialysis before successful thrombin cleavage could be carried out. Cleavage with thrombin at 18°C for 16hrs produced entirely cleaved TFL1. Final purification to remove the thrombin was carried out on a

MonoS column and resulted in approximately 5mg of pure (greater then 95% as visualised by SDS-PAGE), cleaved TFL1 per litre of *E. coli* culture (Figure 6.7b).

6.3.2.2 FT

Approximately a third of FT is expressed in the soluble form allowing for a simple two column purification. The FT was initially purified from other soluble proteins by a native Ni^{2+} affinity column where it elutes as the main protein in the second peak with approximately 0.4M imidazole (Figure 6.8). The FT containing fractions were pooled and the His-tag removed by cleavage with thrombin overnight. Separation of the thrombin and any remaining contaminating proteins was achieved by gel filtration (Figure 6.9) with the elution of the main peak corresponding to a monomeric FT weight. This produced approximately 10mg of pure (greater then 95% as visualised by SDS-PAGE), cleaved FT per litre of culture which was used for crystallisation (Figure 6.10).

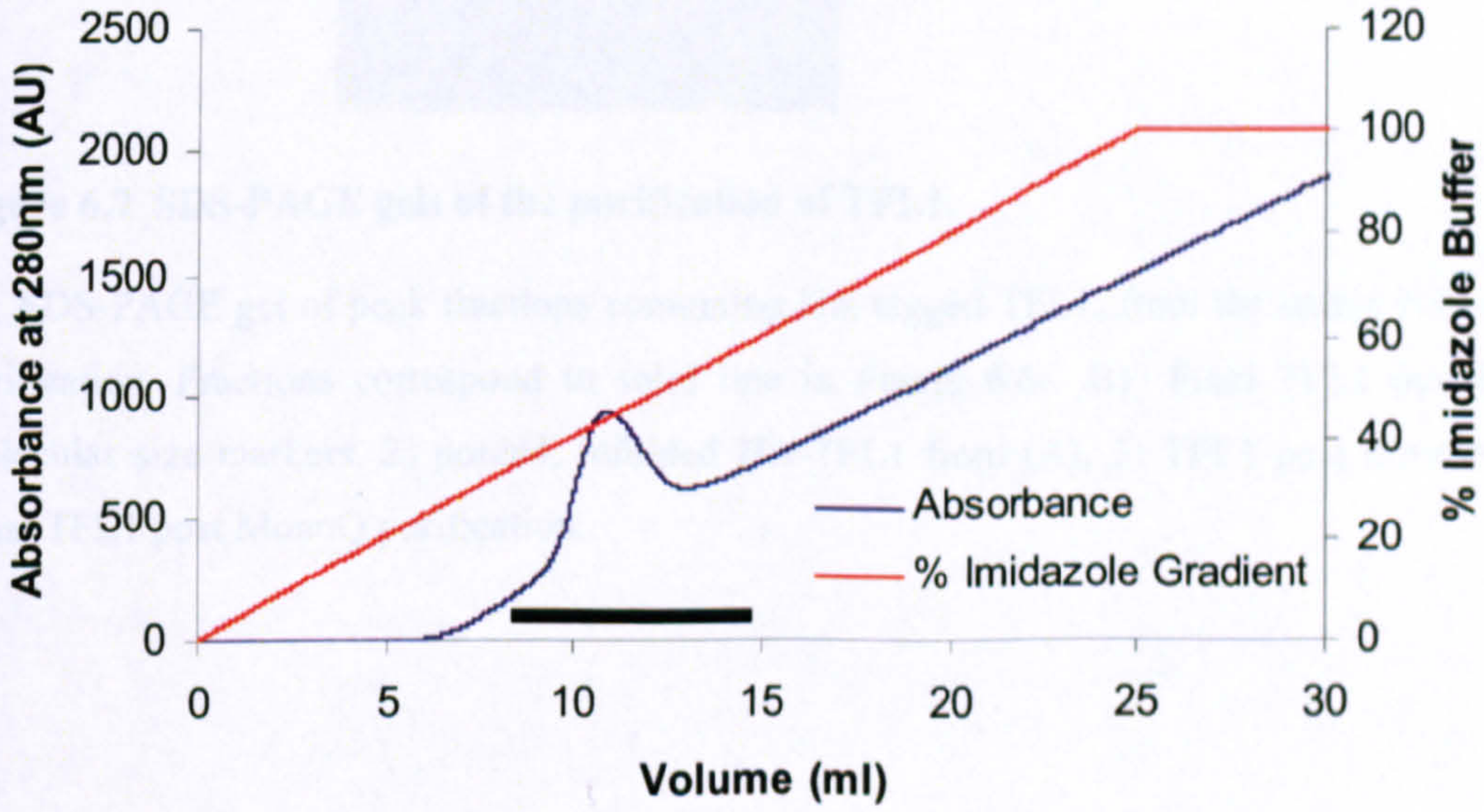


Figure 6.6 Native nickel affinity column chromatography of refolded TFL1.

Profile of refolded TFL1 elution with an imidazole gradient, monitored by UV absorbance at 280nm. Solid line indicates range of fractions run on SDS-PAGE gel in Figure 6.7.

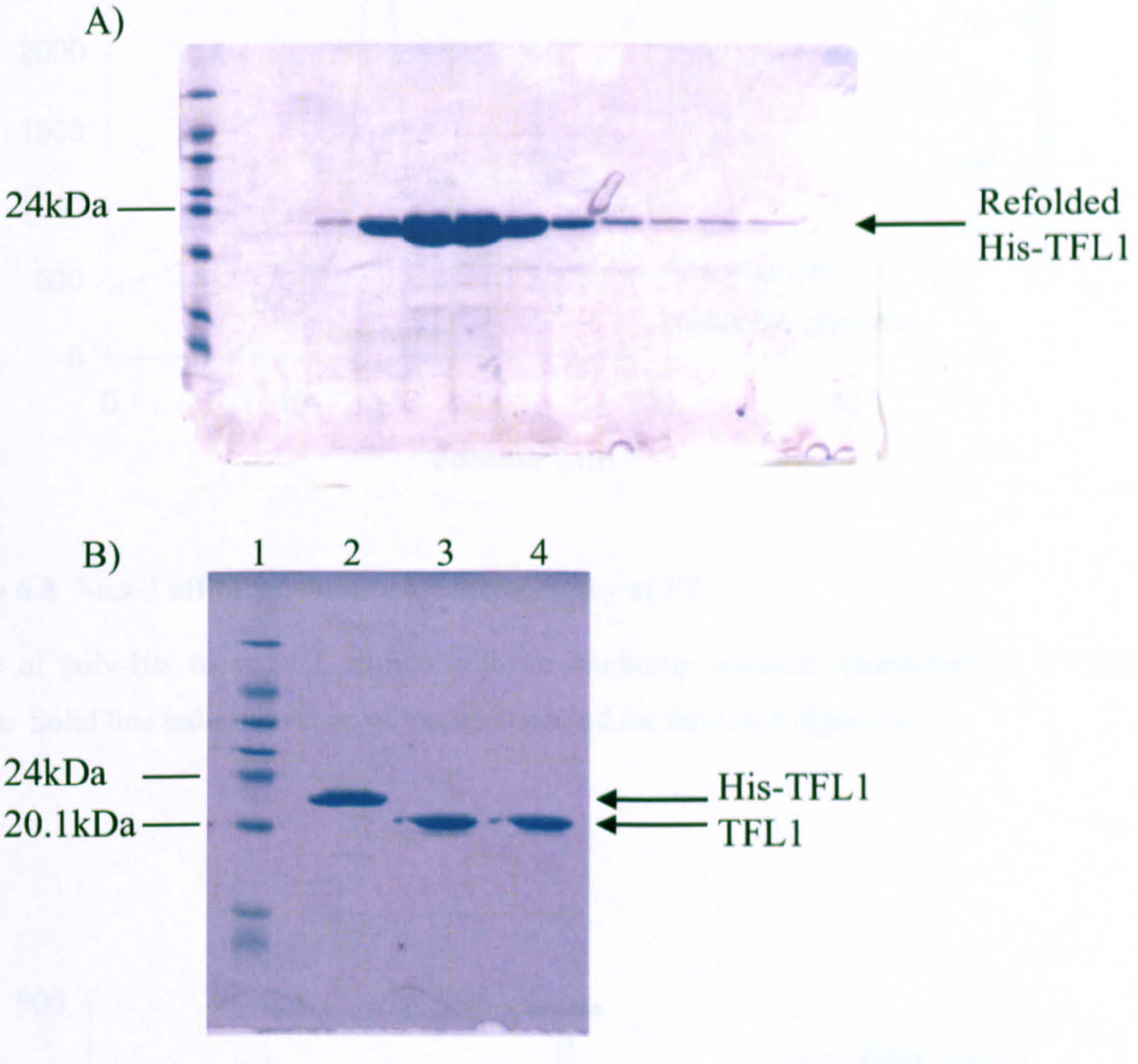


Figure 6.7 SDS-PAGE gels of the purification of TFL1.

A) SDS-PAGE gel of peak fractions containing His tagged TFL1, from the native Ni^{2+}-affinity column purification. Fractions correspond to solid line in Figure 6.6. B) Final TFL1 purification; lane 1) molecular size markers, 2) pooled, refolded His-TFL1 from (A), 3) TFL1 post thrombin cleavage, 4) Final TFL1 post MonoQ purification.

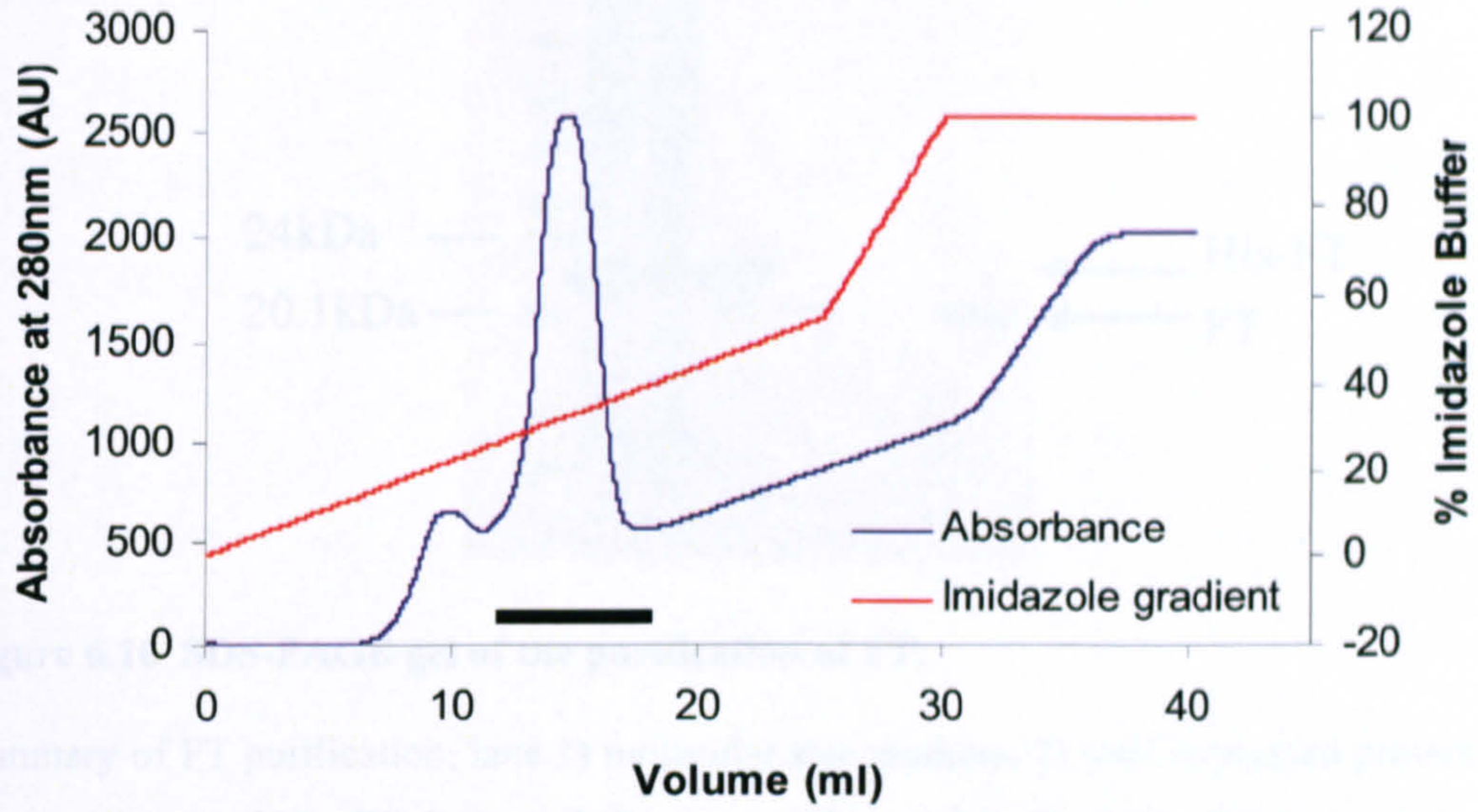


Figure 6.8 Nickel affinity column chromatography of FT.

Profile of poly-His tagged FT elution with an imidazole gradient, monitored by UV absorbance at 280nm. Solid line indicates range of fractions pooled for thrombin digestion.

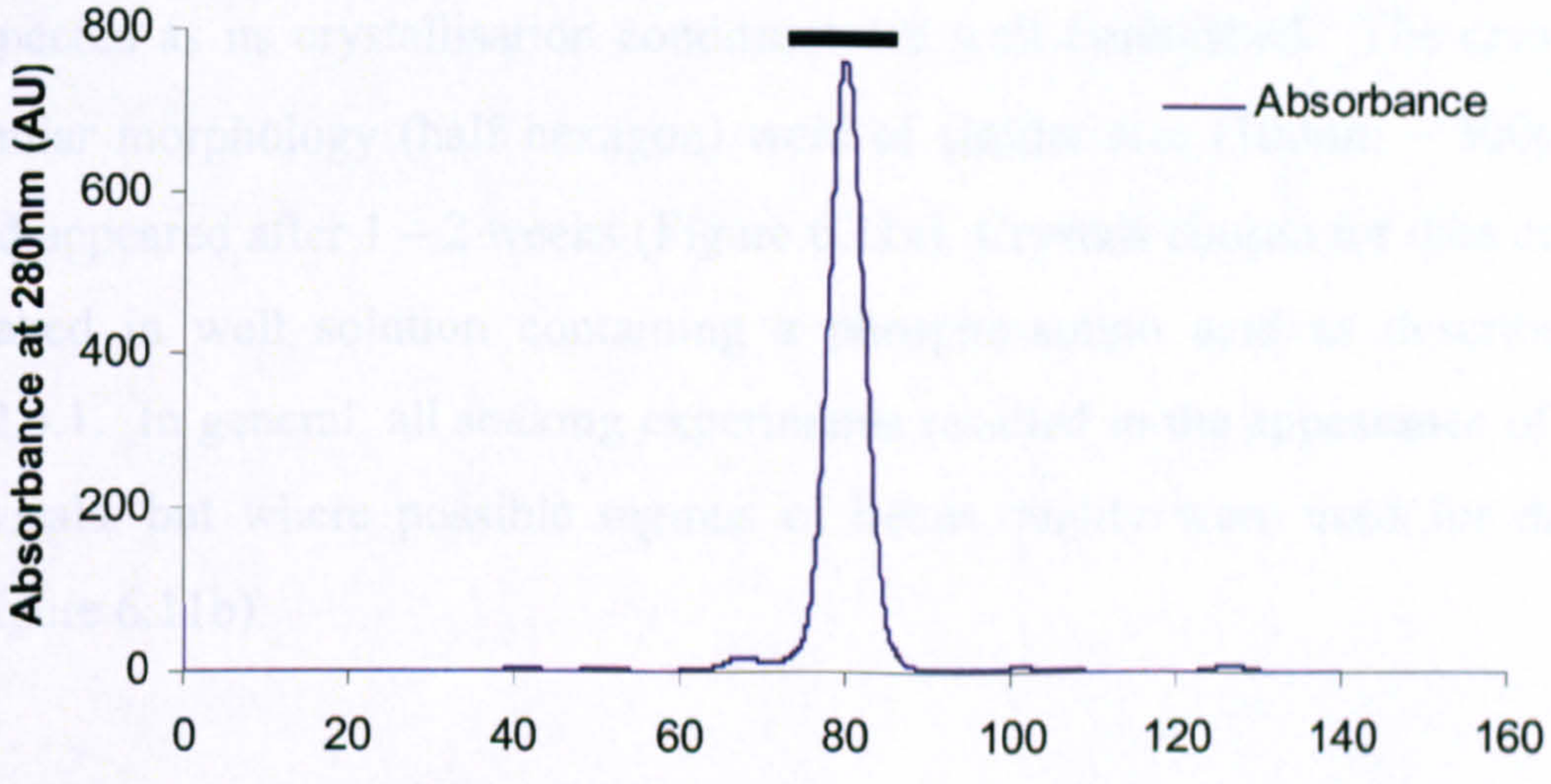


Figure 6.9 S75 gel filtration chromatography of cleaved FT.

Protein elution was monitored by UV absorbance at 280nm. The peak corresponds to monomeric FT and the solid line indicates those fractions combined to provide protein for crystallisation.

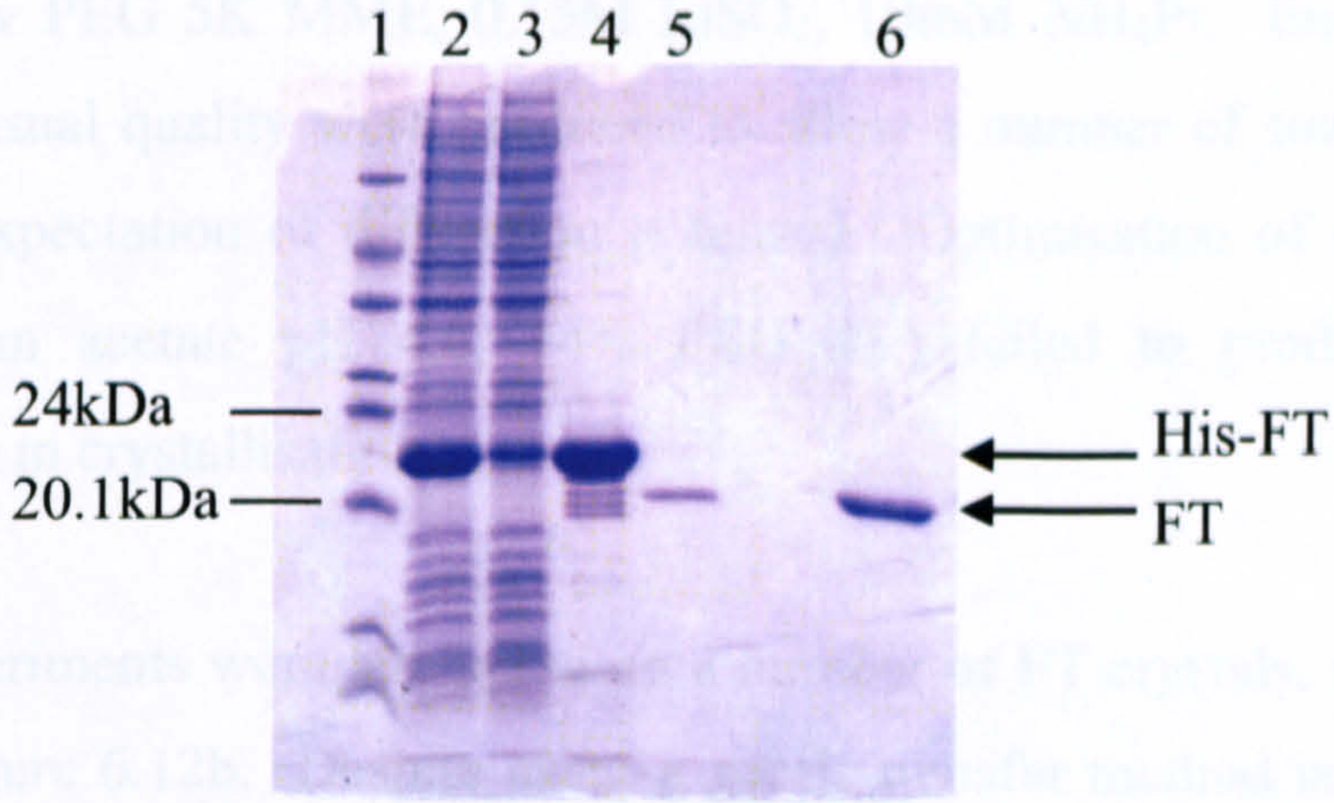


Figure 6.10 SDS-PAGE gel of the purification of FT.

Summary of FT purification; lane 1) molecular size markers, 2) total expressed protein, 3) total soluble protein, 4) pooled His-FT from peak fractions of Figure 6.8, 5) pooled FT post gel filtration, from peak fractions of Figure 6.9, 6) concentrated FT post gel filtration.

6.3.3 Phospho-amino acid soaking experiments

6.3.3.1 TFL1

Good quality crystals of TFL1 grew in a number of the conditions tried. This was expected as its crystallisation conditions are well established. The crystals all had a similar morphology (half hexagon) were of similar size (100μm – 300μm in length) and appeared after 1 – 2 weeks (Figure 6.11a) Crystals chosen for data collection were *soaked in well solution containing a* phospho-amino acid as described in section 6.2.4.1. In general, all soaking experiments resulted in the appearance of cracks in the crystals, but where possible regions of better quality were used for data collection (Figure 6.11b).

6.3.3.2 FT

The crystallisation of FT has been less reproducible and many of the conditions tried resulted in multiple small, thin plates (Figure 6.12a). By using the Hampton additive screen it was possible to achieve a well formed FT crystal (Figure 6.12b). This crystal grew in 38% PEG 5K MME, 0.1M MES pH 7.0, 0.14M $(NH_4)_2SO_4$, 10mM NH_4Pi and 3% v/v glycerol, but was only around 100μm in length and very thin. This proved difficult to reproduce, as did a second reasonable quality apo condition, 0.1M MES pH

7.0, 32% w/v PEG 5K MME, 0.15M $LiSO_4$, 10mM NH_4Pi. Enough FT crystals of acceptable visual quality were produced to allow a number of soaks to be carried out with some expectation of diffraction potential. Optimisation of the novel screen hit (0.1M sodium acetate pH 4.6, 14% PEG 4K) failed to produce any significant improvement in crystallisation.

Soaking experiments were attempted on a number of FT crystals, including the crystal shown in Figure 6.12b. Despite using a gentle transfer method into the soak solution many soaking attempts resulted in the crystals breaking apart, dissolving or severely cracking. Of those crystals where diffraction was tested only one produced analysable data, a 24hr soak with 0.1M p-Ser. As FT crystals are known to be of poor quality (personal communication Dr D. Miller, University of Bristol) it is unclear if the lack of collectable data sets was due to the crystals being of low quality before soaking or if the soaking caused (or enhanced) crystal disorder.

a. b.

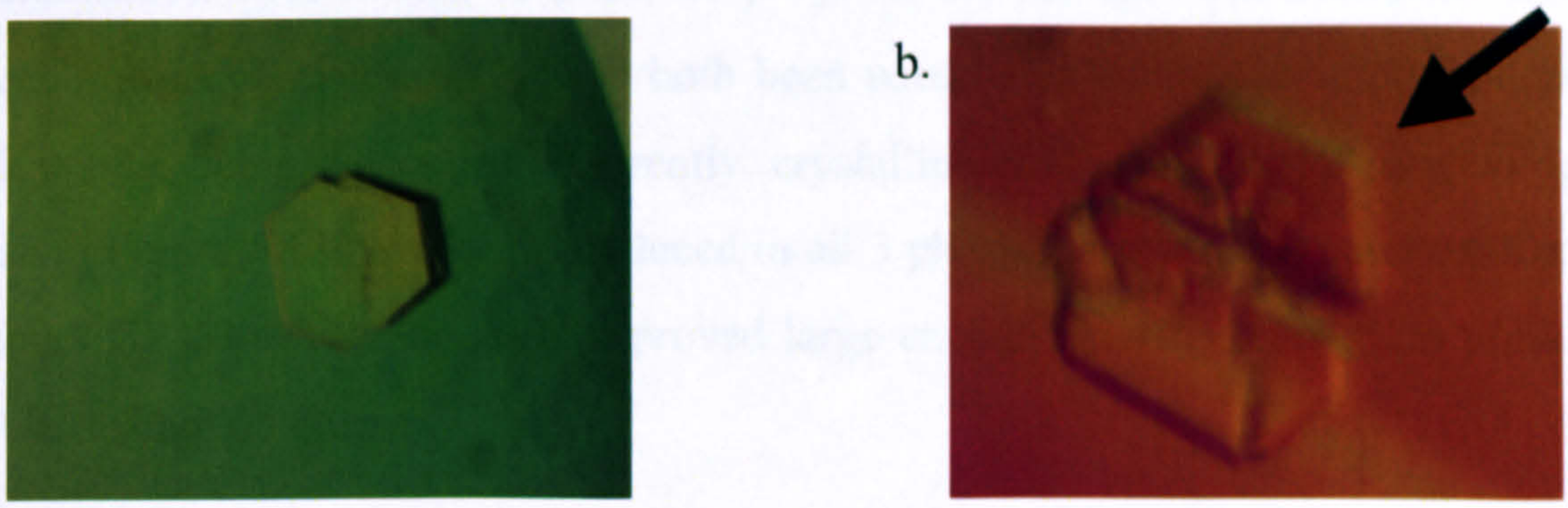


Figure 6.11 Example TFL1 crystals.

A) Apo TFL1 crystal. B) Same crystal as in (A) after 72hrs soaking with phospho-amino acid (p-Tyr). Arrow indicates region used for data collection.

a. b.

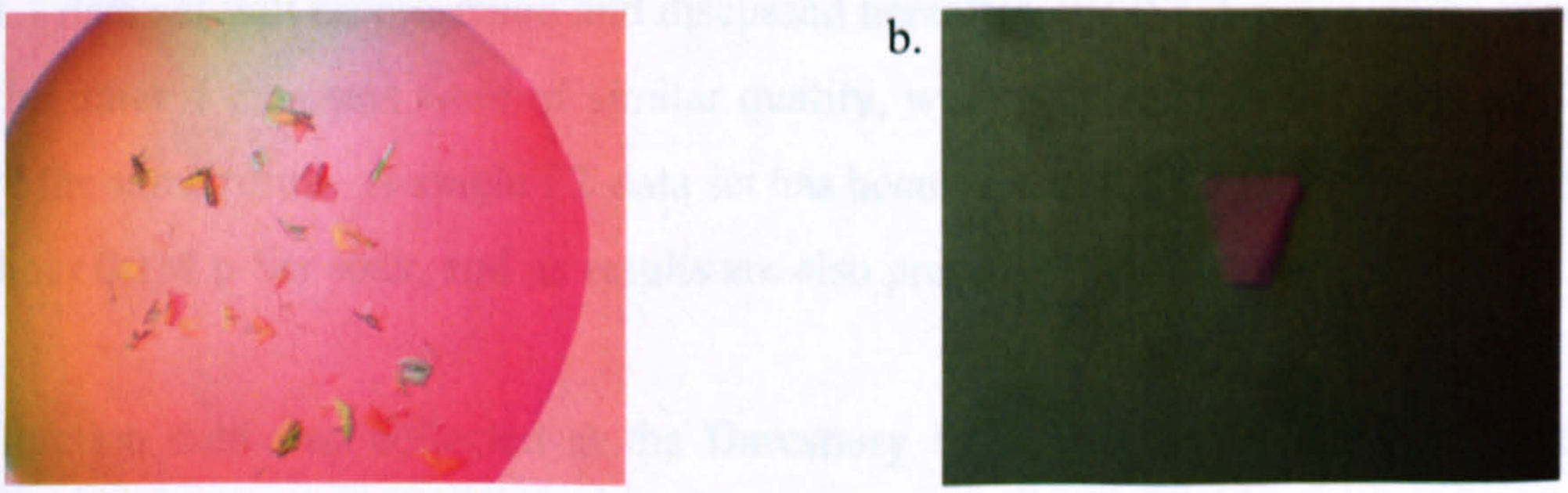


Figure 6.12 Example FT apo crystals.

A) Standard morphology of FT crystals produced. B) Example FT crystal produced using additive screen.

6.3.4 Phospho-amino acid co-crystallisation

A number of potential conditions have been identified for the production of FT/phospho-amino acid co-crystals. The MDL Screens I and II were carried out with 50mM p-Ser and produced small crystalline plates after a month in 0.1M Na Citrate pH 5.6, 30% w/w PEG 4K, 0.2M ammonium acetate. The JBScreen 3 is a PEG 4K based commercial screen and was undertaken with 15mM p-Ser, resulting in small crystalline plates in the same condition, but crystal growth only took one week. In addition JBScreen 3 showed phase separation at 25% w/w PEG 4K with the same buffer and salt, providing an indication of the low range of PEG 4K concentration to use in screening. Crystal optimisation was carried out with 15mM p-Ser, p-Tyr and p-Thr, 26% - 38% PEG 4K, 0.16M – 0.22M ammonium acetate in 0.1M Na Citrate pH 5.6. Additional screening was done by taking the best apo condition (0.1M MES pH 7.0, 32% w/v PEG 5K MME, 0.15M $LiSO_4$, 10mM NH_4Pi) and using this as a basis for co-crystallisation with 30mM of p-Ser or p-Tyr in the set up. The Hampton additive screen and incubation at 4°C have both been used in an attempt to improve crystals, with neither being successful. Currently, crystalline plates of a similar morphology to those in Figure 6.13 have been produced in all 3 phospho-amino acid co-crystallisation (Table 6.3). However, none have proved large enough or stable enough to produce a usable diffraction pattern.

6.3.5 Data collection

In total 5 data sets have been collected for TFL1 phospho-amino acid soaking experiments, 2 each for p-Tyr, p-Ser and 1 for p-Thr (Table 6.4). However, only one TFL1 data set will be presented and discussed hereafter, the 0.1M p-Tyr 72 hour soak, as the other 4 data sets were of similar quality, were processed in the same way and gave the same result. A single FT data set has been successfully collected and solved, a 24 hour 0.1M p-Ser soak, and its results are also presented below.

Diffraction data was collected at the Daresbury SRS, Warrington under cryocooled conditions (100K). Test diffraction images were used to optimise exposure time and detector distance. Additionally, the strategy program within HKL2000 was used to

determine how best to collect complete data sets. The collection parameters are given below in Table 6.5 and example diffraction images are given in Figure 6.14.

Phospho-amino acid (15mM)	**Best condition**
p-Ser	0.1M Sodium Citrate pH 5.6, 34% PEG 4K, 0.16M Ammonium acetate
p-Tyr	0.1M Sodium Citrate pH 5.6, 38% PEG 4K, 0.16M Ammonium acetate
p-Thr	0.1M Sodium Citrate pH 5.6, 38% PEG 4K, 0.18M Ammonium acetate

Table 6.3 Best determined co-crystallisation condition for FT with each phospho-amino acid.

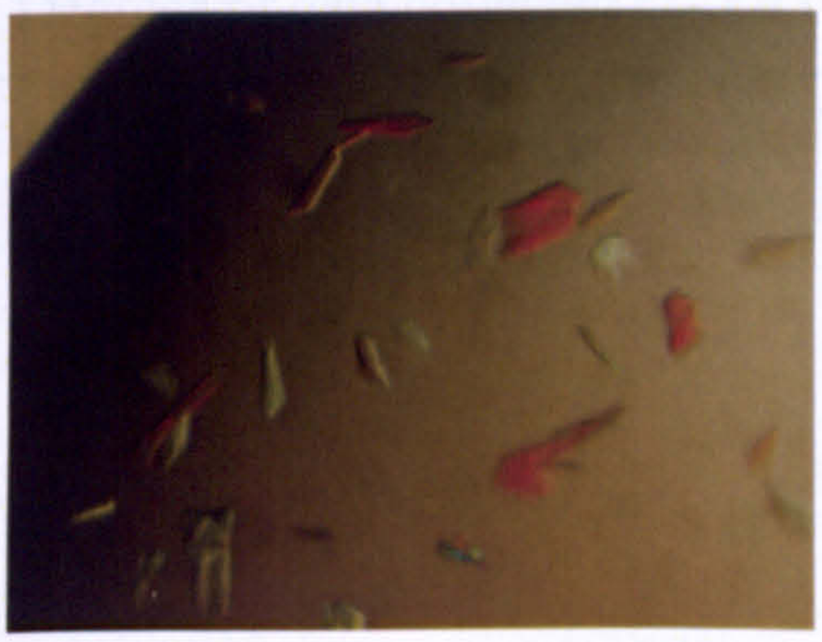


Figure 6.13 Example FT/p-Ser co-crystals.

Co-crystals produced in the p-Ser condition give in Table 6.3. These did not produce usable diffraction.

Phospho-amino acid (0.1M)	***Soak length***
p-Tyr	72 hours
	1 week
p-Ser	3 weeks
	2 months
p-Thr	2 months

Table 6.4 Length of phospho-amino acid soaking times used on TFL1 crystals.

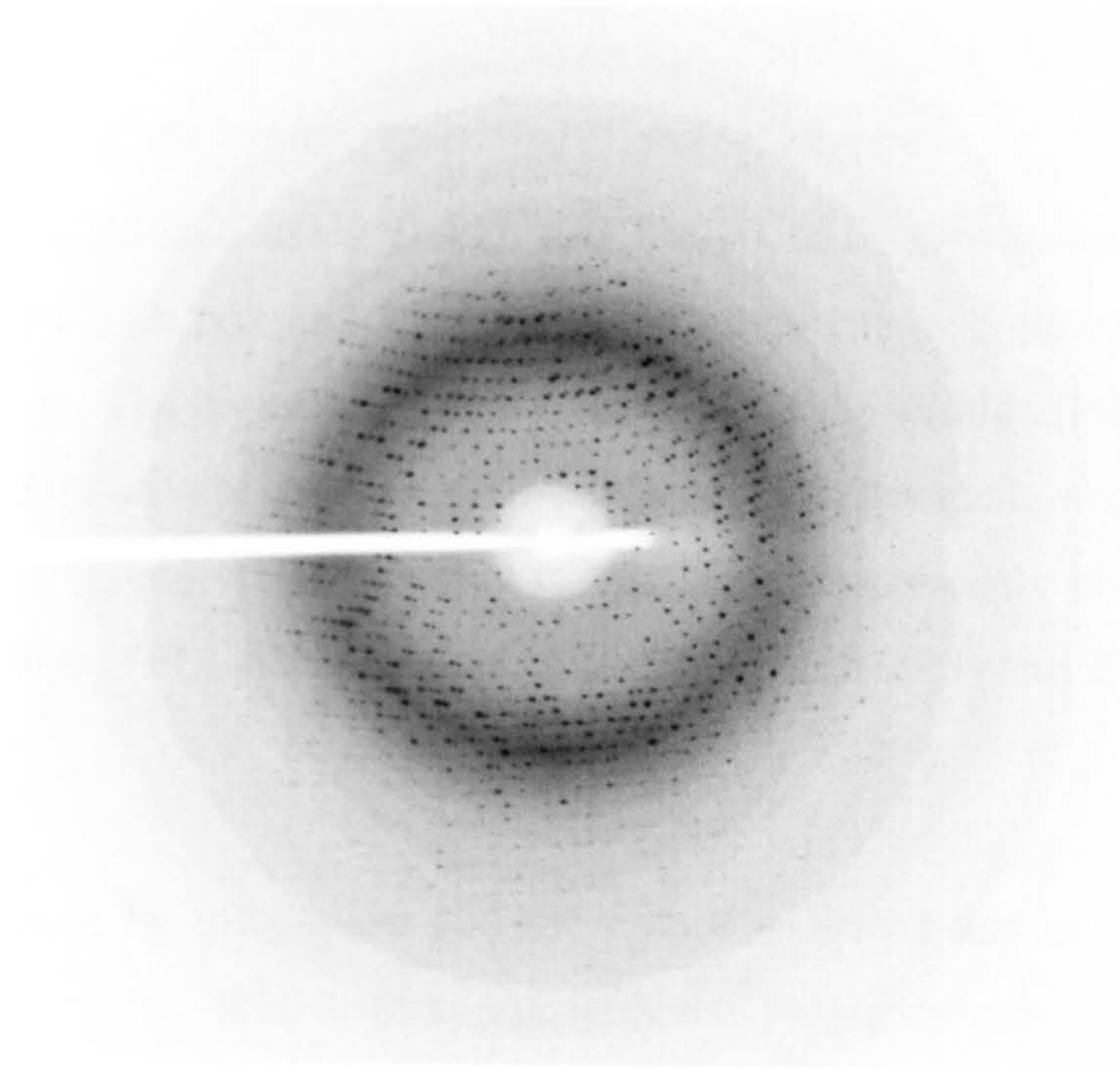
The length of time each data collected TFL1 apo crystal was soaked in 0.1M phospho-amino acid prior to data collection (or freezing).

	TFL1	FT
Crystal Conditions	10mg.ml^{-1}, 0.1M MES pH 6.5, 28% PEG 5K MME, 0.3M $(NH_4)_2SO_4$	10mg.ml^{-1}, 0.1M MES pH 7.0, 38% PEG 5K MME, 0.14M $(NH_4)_2SO_4$, 10mM NH_4Pi and 3% v/v glycerol
Soak	0.1M p-Tyr, 72 hours	0.1M p-Ser, 24 hours
Cryoprotectant	20% glycerol, 0.1M p-Tyr	5% glycerol, 0.1M p-Ser
Station/Wavelength	10.1/1.488	14.1/1.488
Detector	MAR165 CCD detector	Quantum 4 CCD detector
Exposure time (sec)	7	45
Oscillation per frame (°)	0.5	1
Detector distance (mm)	100	120
Degrees of data collected (°)	239	327

Table 6.5 Data collection parameters for example TFL1 and FT data sets.

All data collection on FT and TFL1 was undertaken at the Daresbury synchrotron, Warrington.

A)



B)

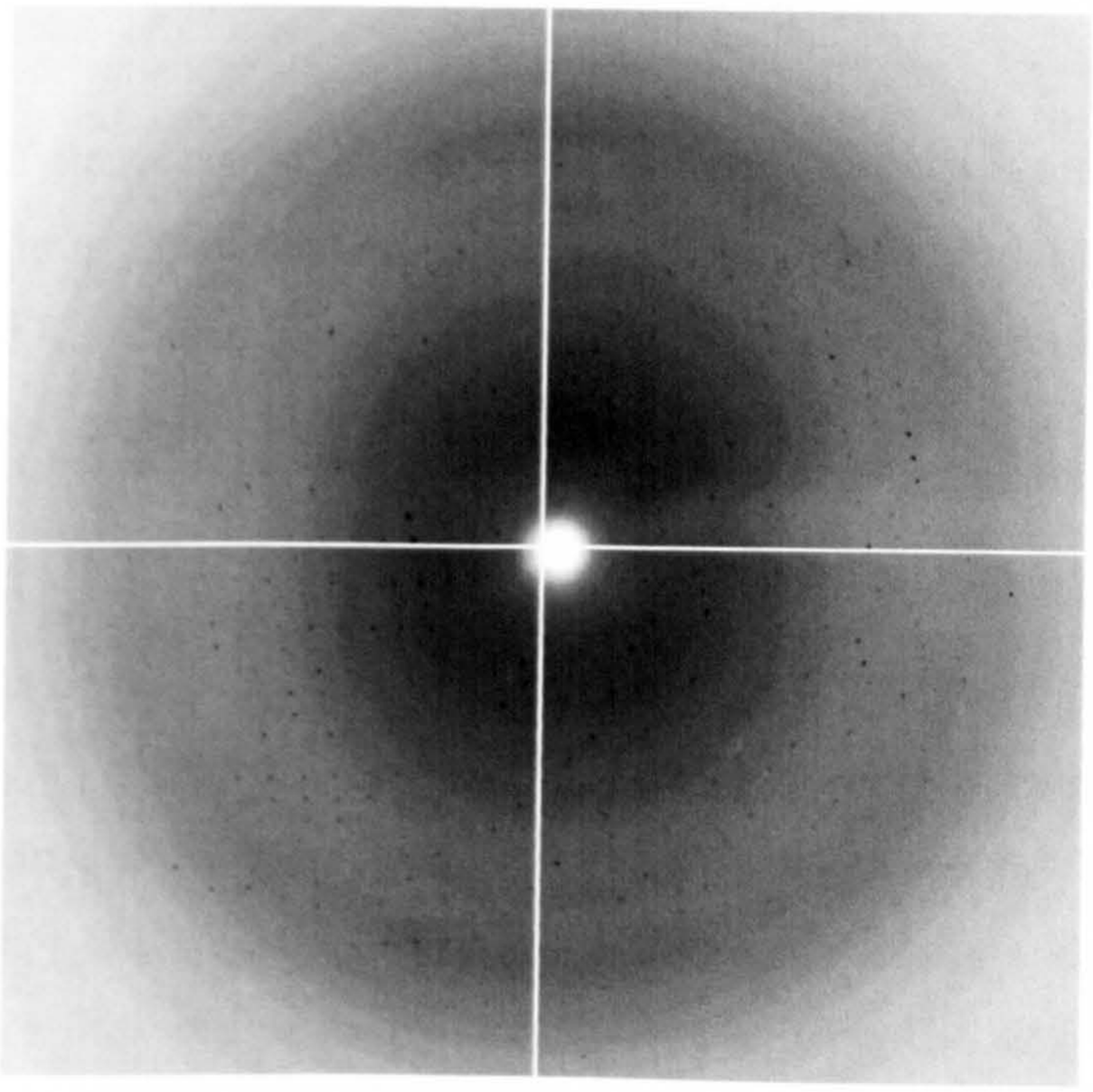


Figure 6.14 Images of diffraction pattern from TFL1 and FT crystals.

A) Example diffraction pattern from TFL1 crystal, B) Example diffraction pattern from FT crystal. This shows the poor diffraction ability of FT crystals in comparison to TFL1. Data collect parameters were as given in Table 6.5.

6.3.6 Data processing, molecular replacement and refinement

Diffraction data was indexed using Denzo (HKL2000 suite, Otwinowski and Minor, 1997), initially in P1 with box and spot size adjustments made, followed by multiple rounds of refinement. The highest symmetry Bravais Lattice with the lowest distortion was then chosen (data not shown) and further rounds of refinement were undertaken. The refined unit cell measurements closely matched those previously published as did the lattice choice, with TFL1 indexed as primitive hexagonal (P3) and FT as primitive triclinic (P1).

Indexing was followed by integration and the data was scaled and merged onto one scale with ScalePack (HKL2000 suite). As with the published structure, TFL1 was scaled in P321. The processing statistics are shown in Table 6.6.

The Matthews coefficient was calculated to determine the number of protein monomers in the asymmetric unit. This showed TFL1 with 2 monomers per AU and a solvent content of 55%, while FT contained 4 monomers and a solvent content of 51% (data not shown) consistent with previous work (Ahn *et al.*, 2006).

Molecular replacement was carried out using Phaser (Storoni, McCoy and Read, 2004). A single protein chain was extracted from the PDB files, 1WKO and 1WKP for TFL1 and FT respectively, to act as the search model. A high quality solution was found for each and examined in Coot (Emsley and Cowtan, 2004). The maps produced by Phaser were of good quality due to the 100% identity of search model and structure. The model required very little alteration, indicating that soaking had had little effect on side-chain and loop position. Any required changes were made followed by rounds of refinement in Refmac5 (Murshudov, Vagin and Dodson, 1997) and subsequent inspection of the resulting density maps. Upon examination of the putative phospho-amino acid binding pocket no density was observed to indicate the presence of the soaked ligand, nor was there evidence for the ligand being present elsewhere in the asymmetric unit. Figure 6.15 shows a view of the proposed binding pocket in both TFL1 and FT. This indicates both the quality of the 2Fo-Fc map for the surrounding residues, and the fact that there is no suggestion of density within the pocket.

	TFL1	FT
Processing		
Space group	P321	P1
Unit cell (Å)	a/b = 108.8, c = 65.0 γ = 120°	a = 53.9, b = 62.5, c = 67.8 α = 74.2°, β = 72.2°, γ = 70.3°
Resolution range (Å)	50 – 1.75 (1.81 – 1.75)	50 – 2.70 (2.80 – 2.70)
Number of unique reflections	44926 (4477)	18788 (1279)
I/σ	36.2 (3.7)	8.4 (1.9)
R_{sym} (%)	8.2 (38.8)	12.6 (42.3)
Completeness (%)	100 (100)	87.1 (59.6)
Redundancy	14.4 (12.7)	3.1 (2.4)
Mosaicity (°)	0.414	2.173
Refinement		
R_{cryst} (%)	24.8	18.3
R_{free} (%)	27.0	29.1

Table 6.6 Final refinement statistics for representative TFL1 and FT data sets.

Values are for the complete dataset, bracketed values represent the highest resolution shell only.

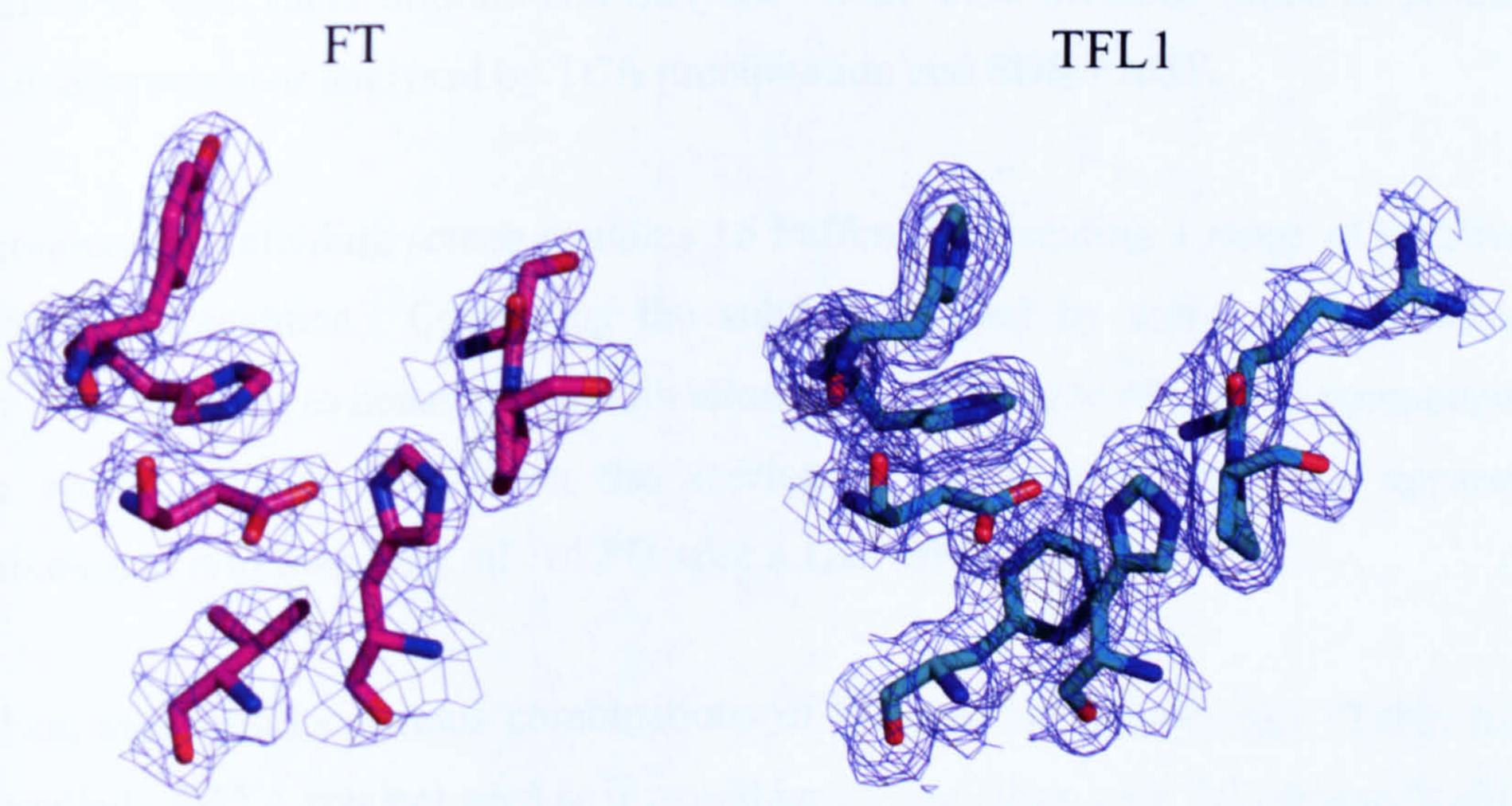


Figure 6.15 Quality of electron density for the pocket forming residues.

The electron density maps (2Fo-Fc) of the residues forming the proposed phospho-amino acid binding pocket for FT and TFL1. This indicates the quality of side-chain electron density and the lack of density within the pocket. The residues shown are identical to those in Figure 6.5. Maps are contoured at 1σ.

6.3.7 FD purification

The expression of FD occurs from the pET-16b vector and attaches an additional 38 amino acids to the N-terminus, including a deca-His tag. Small scale expression tests were carried out across a range of temperatures and IPTG concentrations, however soluble protein was not seen (personal communication Dr T. Joseph-Horne, University of Bristol), leading to the need to purify the protein from inclusion bodies. A denaturing Ni^{2+} affinity column was carried out as for TFL1 and an imidazole gradient resulted in the elution of reasonably pure FD at very low imidazole concentrations (Figure 6.16 and Figure 6.17). While some contamination remained these samples were deemed pure enough to be useful for refolding experiments.

6.3.8 Refolding attempts

Over time urea can break down and cause the carbamylation of lysine and arginine side chains. As this may effect folding, only fresh preparations of denatured FD were used in all folding experiments. Initial 'basic' refolding buffers were made and refolding attempted by both rapid dilution and dialysis. Both these methods failed to produce any soluble protein as analysed by TCA precipitation and SDS-PAGE.

The commercial refolding screen contains 15 buffers representing a range of additives and their combinations. Comparing the solublising (and by conjecture, refolding) ability of each buffer to denatured protein allows the importance of various components to be scored. Table 6.7 shows the scoring of components for their apparent contribution to refolding $1 mg.ml^{-1}$ of FD after a 1:20 dilution.

Based on these results various combinations of the positive components (Table 6.7) were trialled. EDTA was not used as it would be incompatible with further purification by Ni^{2+} affinity. Additionally, the detergent CHAPS was used to replace Triton X-100 as this can prove troublesome for crystallisation experiments. A final buffer of 50mM MES pH 6.0, 250mM NaCl, 10mM KCl, 0.5M L-arginine, 0.5% w/v CHAPS proved successful at retaining FD solubility after a 1:20 dilution. However, upon either buffer exchange by centrifugation (Vivaspin, 10kDa cutoff) or dialysis into 50mM MES pH 6.0, 250mM NaCl, the FD precipitated. This suggests that incomplete or incorrect

folded versions of FD were being retained in the supernatant by the general solubilising ability of L-arginine and CHAPS rather then actual successful refolding.

Additionally on-column refolding was attempted. The method used here, "artificial chaperone-assisted refolding" was developed by (Oganesyan, Kim and Kim, 2005). This uses slow replacement of the denaturant (urea) with a detergent-containing buffer (in this case Triton X-100) to prevent aggregation, followed by removal of the detergent with β-cyclodextrin. However, this method proved ineffective in the refolding of full-length FD and resulted in its elution as an insoluble aggregate after the removal of urea.

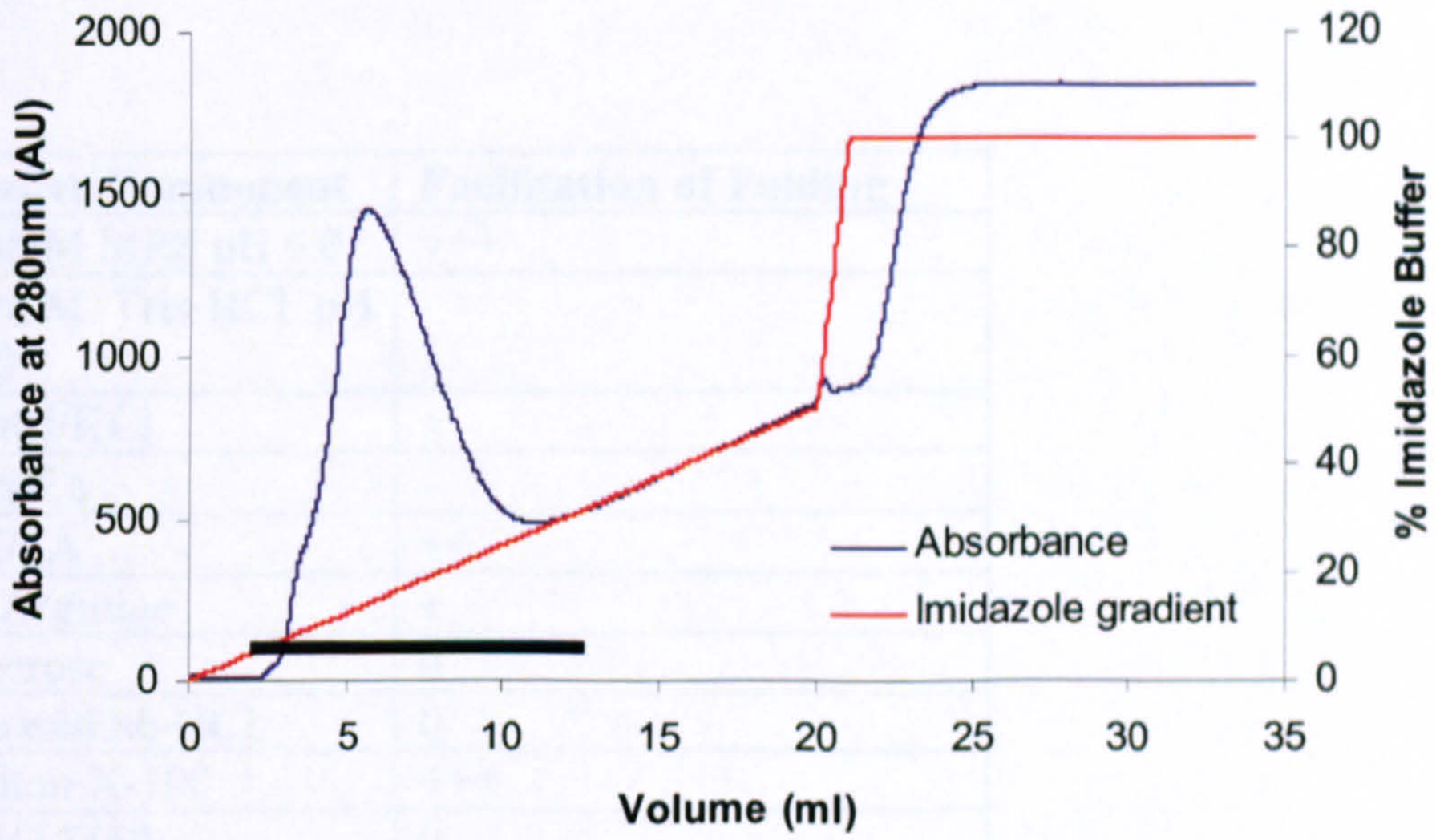


Figure 6.16 Denaturing nickel affinity column chromatography of FD.

The nickel column was run in the presence of 8M urea and FD elution monitored by UV absorbance at 280nm. Solid line indicates the range of fractions ran on the SDS-PAGE gel in Figure 6.17.

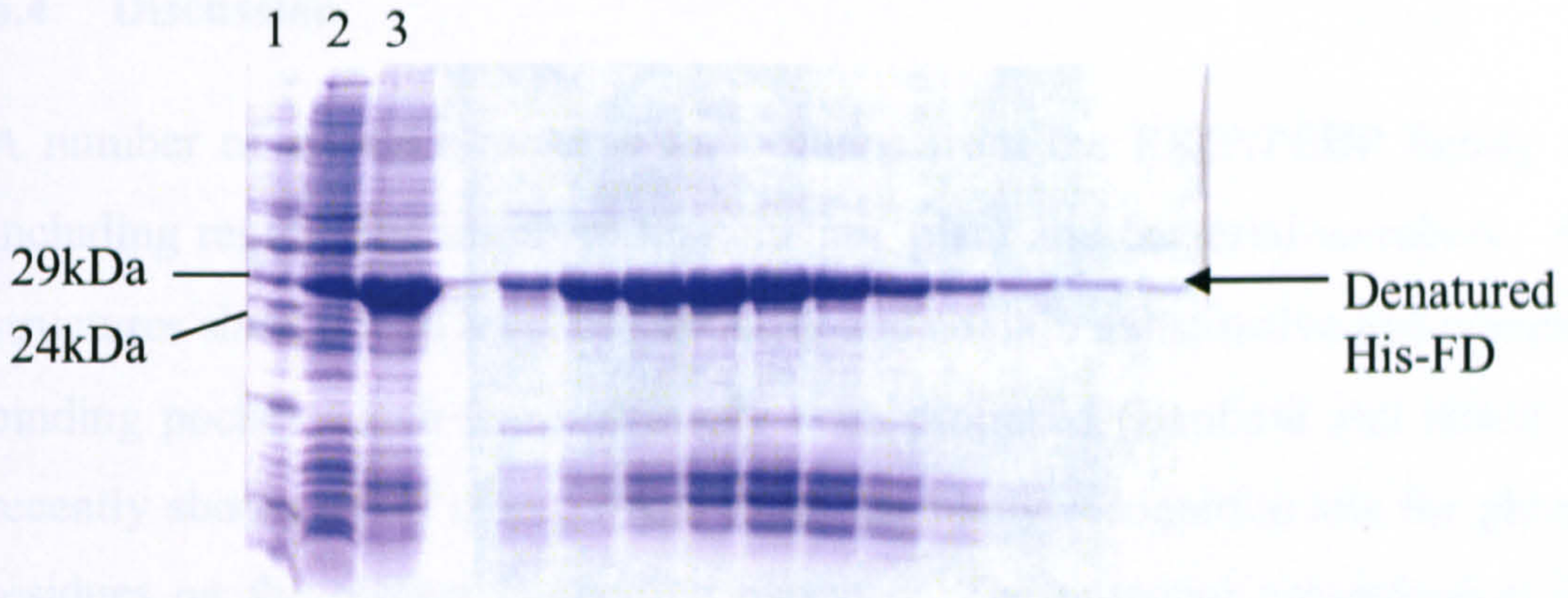


Figure 6.17 SDS-PAGE gel of the denatured purification of His-FD

Lanes 1) molecular size markers, 2) total expressed protein, 3) total insoluble protein. The remaining lanes correspond to the region indicated by the solid line in Figure 6.16. Those fractions containing the majority of FD were pooled for refolding experiments.

Buffer Component	Facilitation of Folding
50mM MES pH 6.0	+++
50mM Tris-HCl pH 8.0	-
NaCl/KCl	+
Mg/Ca	-
EDTA	++
L-Arginine	+
Sucrose	0
Guanidine-HCL	0
Triton X-100	+++
PEG 3550	0
DTT	0
GSH/ GSSH	0

Table 6.7 MDL refolding screen components and their effect on FD folding.

0 = no effect, - = negative effect., + = some positive effect, ++ = reasonable positive effect, +++ = clear positive effect.

6.4 Discussion

A number of crystal structures for proteins from the RKIP/PEBP family are known, including representatives from mammalian, plant and bacterial members. All of these structures show a high level of similarity and contain a distinctive and conserved anion-binding pocket which has previously been proposed (Banfield and Brady, 2000) and recently shown (Rath *et al.*, 2008) to be the likely recognition site for phosphorylated residues on the respective binding partners. The potential physiological interactions facilitated by this protein family remain structurally uncharacterised, with the exception of the structure of human RKIP bound to phospho-tyrosine (Simister, 2004). This led us to investigate if a similar interaction could be observed with the plant proteins TFL1 and FT. The results of this work are described here, with a focus on possible reasons for the lack of success and suggestions for future directions that could be followed. The purification and structural details of TFL1 and FT will not be discussed as they are identical to those published elsewhere (Miller, 2004; Ahn *et al.*, 2006).

6.4.1 Soaking and co-crystallisation experiments

A large constraint on both the soaking and co-crystallisation experiments was the lack of clear knowledge of which (if any) phospho-amino acid binds to the pocket. Attempts were made to study the potential phospho-amino acid interactions via ITC. However it proved impossible to solubilise the protein and phospho-amino acid in identical buffers, a requirement for ITC. Although the related protein (SP) from tomato has been reported to interact with a phospho-serine in at least some of its binding partners (Pnueli *et al.*, 2001), it was not until after the FT/TFL1 work had been initiated that a potential interaction dependent on either a (likely phosphorylated) threonine or serine was identified between FT and FD (Abe *et al.*, 2005). Consequently, the later co-crystallisation work focused on this interaction, assisted by the fact that FT is simple to purify in the large quantities needed for extensive screening. If phospho-serine (p-Ser) is the appropriate ligand for FT it is disappointing that the crystallographic data produced from a 24hr p-Ser soak into a FT crystal (sections 6.3.3.2 and 6.3.6) did not contain any discernible density for this ligand.

All soaking experiments led to deterioration in the visual quality of both TFL1 and FT crystals, although only FT crystals were seen to totally disintegrate over time while

TFL1 crystals remained reasonably intact for at least 2 months. While care was taken to minimise the change in buffer condition upon the addition of the phospho-amino acid it appears that, due to the known delicate nature of apo FT crystals (Miller, 2004), this slight change in conditions has had a significant deleterious effect. Whether this is due to a change in buffer conditions or a ligand interaction disrupting crystal packing is unclear.

Successfully soaking a ligand into apo crystals is dependent on a number of factors, including the retention of important crystal lattice interactions, the affinity of the protein:ligand interaction and, importantly, the presence of solvent channels allowing the ligand access to the binding site. Examination of the TFL1 crystal packing indicates that one of the two monomers in the asymmetric unit obstructs access to the opposing binding pocket. However, that monomer itself retains a pocket apparently accessible to solvent. The crystal packing and interactions between the 4 monomers within the asymmetric unit of the FT crystal allow all 4 binding pockets to face solvent channels. This would suggest that the lack of electron density for a phospho-amino acid ligand is not due to its inability to gain access to the interaction site. The RKIP:p-Tyr structure resulted from a 72hr, 0.1M soak, with shorter times being unsuccessful (Simister, 2004). While this time frame was achieved for the TFL1 soak experiments it could not be reproduced with FT due to crystal deterioration. Consequently the absence of ligand in the FT binding pocket may be the result of an insufficient soak time, although this does not explain its absence in TFL1.

The use of co-crystallisation avoids the problems of ligand access as it is present prior to crystal lattice formation, assuming adequate binding affinity. The 1.4 pH unit shift seen between the apo (pH 7.0) and initial co-crystallisation (pH 5.6) conditions (Table 6.2 and Table 6.3) might be indicative of an altered lattice and, potentially, the presence of the ligand. Alternatively the phospho-amino acid could merely be acting as an additive within the buffer and not be incorporated into the crystal. However, this seems unlikely given the relatively low (15mM) concentration of the ligand. Further work need to be undertaken on these co-crystallisation conditions, particularly to establish conditions for the formation of a stable protein-ligand complex, with the aim of producing diffraction quality crystals.

It is possible that the interaction seen in the RKIP:p-Tyr structure is a crystallographic artifact and the conditions that favored this simply have not been reproduced with TFL1 or FT. There is however mounting experimental evidence that the pocket on this family of proteins does indeed bind phosphorylated residues (Pnueli *et al.*, 2001; Abe *et al.*, 2005; Rath *et al.*, 2008), and it seems more likely that the lack of ligand within the soak experiments indicates a very low affinity interaction.

6.4.2 Expression of FD

With the subsequent identification of the potential binding partner, FD (Abe *et al.*, 2005; Wigge *et al.*, 2005), attempts were made to produce recombinant FD to facilitate investigation of the binding interaction. However no soluble full-length protein could be produced. The difficulty of obtaining folded (or in refolding) FD may be due to the large N-terminal extension from the vector (38 amino acids). Alternatively, the large serine and proline rich native N-terminus sequence of FD (residues 1 - 207) may require particular conditions to assist folding or may be intrinsically unstructured. Co-expression with chaperone proteins or expression in a yeast host may be profitable avenues for further investigation of the full length form. Work has also been undertaken by others in the host laboratory on truncated forms of FD, retaining only the C-terminal leucine zipper domain (residues 208 – 285).

6.4.3 Future work

While the extensive phospho-amino acid soaking experiments undertaken during the course of this work failed to produce a protein:ligand complex structure, the production of crystals under co-crystallisation conditions, although of insufficient quality, provide a starting point for future screening work. This co-crystallisation work needs to be extending to include TFL1 which, due to time constraints and the priority placed on other projects, could not be achieved during the course of this thesis. Additional biophysical techniques such as surface plasmon resonace (Biocore) could be utilised to investigate phospho-amino binding. It would be important to confirm the identity of the region of FD to which FT binds, which may allow the co-crystallisation of FT with an appropriate phospho-peptide constituting the proposed binding region. The use of a peptide, rather then a free amino acid, may increase the affinity of the interaction sufficiently to allow crystallography to be successful. This method is currently in the

early stages of being employed with hRKIP (personal communication Dr N. Burton, University of Bristol). Recently, soluble expression of the C-terminus leucine zipper of FD (residues 208 - 285) has been achieved (personal communications Dr T. Joseph-Horne, University of Bristol). This region contains a potential phosphorylation site for CDPK, Thr282, that when mutated abolishes the FT:FD interaction (Abe *et al.*, 2005). While not possible within the time frame of this thesis, the nature of this interaction needs to be further investigated, although it may not constitute a site of direct interaction. Various methodologies could be employed including size exclusion chromatography, analytical ultra centrifugation (AUC), ITC and, if confirmed, crystallography. Other potential FD phosphorlyation sites (many of which are in the unstructured N-terminal region; Jakoby *et al.*, 2002) also need to be investigated to ascertain their role, if any, in the interaction with FT/TFL1.

Chapter 7 General Conclusions

The different proteins investigated throughout this thesis are involved in a variety of cellular processes, from metabolism to trafficking and signaling. Despite this diversity, they are linked by a common theme of the structural mechanisms by which ligand specificity is determined.

X-ray crystallography is a powerful technique for the determination of protein structures and consequently, their interactions with substrates and ligands. The work undertaken during this thesis was aimed at various points along the pathway by which protein crystal structures are determined. These ranged from enabling protein expression studies at one end to the elucidation of structural complexes at the other, with all of these steps being implemented in the study of human GRHPR.

The production of a structure of the human GRHPR enzyme in complex with both its coenzyme (NADPH) and a substrate molecule (D-glycerate) has shown how substrate specificity is achieved within the GRHPR family of enzymes (Chapters 3 and 4). The elegant use of a steric gate to block unwanted substrates such as pyruvate, combined with the deployment of hydrogen bonding potential to allow specific, larger substrates (hydroxypyruvate) to overcome this restriction is a mechanism which could not have been elucidated without a ternary structure. Additionally, this structure has also provided the first structural proof for the manner by which the entire D-2-hydroxyacid dehydrogenase protein family are likely to bind their substrates.

Protein structures are also very important in the process of drug design. As the malfunction of human GRHPR ultimately leads to primary hyperoxaluria type 2 (PH2) disorder, the structures produced in this thesis may assist in the development of *therapeutics for the treatment* of this condition. The structures have already provided a basis from which the reduction in GRHPR *activity associated with* specific PH2-causing missense mutations can be explained, a significant first *step* when contemplating drug design. As some of these changes appear to lead to destabilisation of the enzyme structure, a possible route might be to develop compounds that could

enhance the GRHPR structure. This concept of molecular chaperones is already being explored for mutated forms of AGT, the cause of PH1. A limitation is that the rarity of PH2 disease questions the economic viability of developing compounds that would find only limited application. As discussed in Chapter 4, a more realistic approach may be to moderate the activity of L-LDH in these individuals instead, 'piggy backing' on existing campaigns aiming to develop specific L-LDH inhibitors for the treatment of cancers. The success of this strategy requires a comprehensive understanding of the labyrinthine interactions of the many metabolic pathways to which L-LDH contributes, a challenge that needs to be addressed via a 'systems biology' approach.

One of the major limitations of X-ray crystallography is the requirement for well-ordered protein crystals from which to collect data. The successful packing of a protein into a crystal requires the formation of regular interactions produced by a highly consistent molecular conformation. With proteins, this is often difficult to achieve because of inherent molecular flexibility, the presence of multiple domains or regions of unstructured sequence, and sample heterogeneity. As demonstrated in this thesis for sorting nexin 1, even producing a large multi-domain protein in sufficient quantities for crystallisation studies can be troublesome (Chapter 5). Often the best approach is to truncate a large protein into smaller, biologically relevant sections. Hopefully by producing a compact domain the crystallisation potential can be enhanced and (some) structural questions answered. The expression and purification procedures established in this thesis for the separate PX and Bar domains of sorting nexin 1 (Chapter 5) have brought the structural understanding of the phosphoinositide and membrane specificity of this protein one step closer, with initial crystallisation trials on these domains still proceeding. Structural studies of larger segments of SNX1 are desirable and potential ways to progress this work have been proposed. The requirement for "packable" protein in crystallography can also result in unexpected conformations being selected due to their crystallisation ability. The partially closed conformation of the apo GRHPR is an example of this and indicates the importance of using knowledge from other, related structures (where possible) in the interpretation of features within a crystal packing environment.

Although protein structures are undeniably important, they provide limited information on the complex interplay of the many proteins present within the cell milieu. The

ability of both GRHPR and its cytosolic companion L-LDH to utilise the same substrates (glyoxylate and hydroxypyruvate), one (GRHPR) producing a desirable product whereas the other (L-LDH) products that are potentially pathological (hyperoxaluria and L-glycerate aciduria), has only been partially examined. It was anticipated that GRHPR would demonstrate a clear kinetic advantage over L-LDH with these substrates under physiological conditions: this, surprisingly, proved not to be the case (Chapter 4). This has led to the proposal that it is the cellular conditions themselves that distinguish between these pathways, with the concentration of other substrates (such and lactate) and the ratios of reduced and oxidized co-enzymes that are likely to dictate the effective operating preferences of these enzymes in normal glyoxylate (and hydroxypyruvate) metabolism.

The work undertaken during this thesis on the interactions of two plant RKIP/PEBP proteins, TFL1 and FT, with their potential phosphorylated ligands has shown that even when the crystallisation and structure of a protein has been established, it is not always a simple step to move to higher level complexes (Chapter 6). Neither soaking the expected small molecule ligand into preformed crystals nor including the ligand prior to crystallisation (co-crystallisation) successfully produced the desired complex structures. Although not successful, this work has allowed the acquisition of experience with these techniques that might be more successfully deployed on other systems in the future.

The variety of ways by which proteins achieve specific binding to ligands or substrates is immense. Looking directly at the molecular structures of such complexes remains the best way to understand these key interactions. Despite its limitations, such as those encountered in these studies, X-ray crystallography remains the most widespread and best technique to determine molecular structures. The advances that have been made in the projects detailed in this thesis would not have been possible without the successful development and application of this important method.

Appendix 1

Publication arising from this thesis;

Mdluli K, Booth MP, Brady RL, and Rumsby G. (2005). A preliminary account of the properties of recombinant human glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism. *Biochim Biophy Acta,* **1753**, 209-16.

Booth MP, Conners R, Rumsby G, and Brady RL. (2006). Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase. *J Mol Biol,* **360**, 178-89.

The GRHPR ternary structure and structure factors have been deposited in the Protein Data Bank (PDB; www.pdb.com), reference code 2GCG.

References

(1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-3.

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K. and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, **309**, 1052-6.

Ahn, J. H., Miller, D., Winter, V. J., Banfield, M. J., Lee, J. H., Yoo, S. Y., Henz, S. R., Brady, R. L. and Weigel, D. (2006). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *Embo J*, **25**, 605-14.

Amaya, I., Ratcliffe, O. J. and Bradley, D. J. (1999). Expression of CENTRORADIALIS (CEN) and CEN-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. *Plant Cell*, **11**, 1405-18.

Antoniou, D., Caratzoulas, S., Kalyanaraman, C., Mincer, J. S. and Schwartz, S. D. (2002). Barrier passage and protein dynamics in enzymatically catalyzed reactions. *Eur J Biochem*, **269**, 3103-12.

Arighi, C. N., Hartnell, L. M., Aguilar, R. C., Haft, C. R. and Bonifacino, J. S. (2004). Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. *J Cell Biol*, **165**, 123-33.

Baker, P. J., Sawa, Y., Shibata, H., Sedelnikova, S. E. and Rice, D. W. (1998). Analysis of the structure and substrate binding of Phormidium lapideum alanine dehydrogenase. *Nat Struct Biol*, **5**, 561-7.

Baker, P. R. S., Cramer, S. D., Kennedy, M., Assimos, D. G. and Holmes, R. P. (2004). Glycolate and glyoxylate metabolism in HepG2 cells. *Am J Physiol - Cell Physiol*, **287**, C1359-C1365.

Banfield, M. J., Barker, J. J., Perry, A. C. and Brady, R. L. (1998). Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. *Structure*, **6**, 1245-54.

Banfield, M. J. and Brady, R. L. (2000). The structure of Antirrhinum centroradialis protein (CEN) suggests a role as a kinase regulator. *J Mol Biol*, **297**, 1159-70.

Bernard, N., Johnsen, K., Holbrook, J. J. and Delcour, J. (1995). D175 discriminates between NADH and NADPH in the coenzyme binding site of Lactobacillus delbrueckii subsp. bulgaricus D-lactate dehydrogenase. *Biochem Biophys Res Commun*, **208**, 895-900.

Bernier, I., Tresca, J. P. and Jolles, P. (1986). Ligand-binding studies with a 23 kDa protein purified from bovine brain cytosol. *Biochim Biophys Acta*, **871**, 19-23.

Bhat, S., Williams, E. L. and Rumsby, G. (2005). Tissue differences in the expression of mutations and polymorphisms in the GRHPR gene and implications for diagnosis of primary hyperoxaluria type 2. *Clin Chem*, **51**, 2423-2425.

Booth, M. P., Conners, R., Rumsby, G. and Brady, R. L. (2006). Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase. *J Mol Biol*, **360**, 178-89.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R. and Coen, E. (1997). Inflorescence commitment and architecture in Arabidopsis. *Science*, **275**, 80-3.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. *Proc Natl Acad Sci U S A*, **99**, 2654-9.

Bravo, J., Karathanassis, D., Pacold, C. M., Pacold, M. E., Ellson, C. D., Anderson, K. E., Butler, P. J., Lavenir, I., Perisic, O., Hawkins, P. T., Stephens, L. and Williams, R. L. (2001). The crystal structure of the PX domain from p40(phox) bound to phosphatidylinositol 3-phosphate. *Mol Cell*, **8**, 829-39.

Brooks, G. A., Dubouchaud, H., Brown, M., Sicurello, J. P. and Butz, C. E. (1999). Role of mitochondrial lactate dehydrogenase and lactate oxidation in the intracellular lactate shuttle. *Proc Natl Acad Sci U S A*, **96**, 1129-34.

Carlton, J. (2004). Examination and Characterisation of Mammalian Sorting Nexins. Biochemistry Department. Bristol, University of Bristol.

Carlton, J., Bujny, M., Peter, B. J., Oorschot, V. M., Rutherford, A., Mellor, H., Klumperman, J., McMahon, H. T. and Cullen, P. J. (2004). Sorting nexin-1 mediates tubular endosome-to-TGN transport through coincidence sensing of high- curvature membranes and 3-phosphoinositides. *Curr Biol*, **14**, 1791-800.

Carlton, J., Bujny, M., Rutherford, A. and Cullen, P. (2005a). Sorting nexins--unifying trends and new perspectives. *Traffic*, **6**, 75-82.

Carlton, J. G., Bujny, M. V., Peter, B. J., Oorschot, V. M., Rutherford, A., Arkell, R. S., Klumperman, J., McMahon, H. T. and Cullen, P. J. (2005b). Sorting nexin-2 is associated with tubular elements of the early endosome, but is not essential for retromer-mediated endosome-to-TGN transport. *J Cell Sci*, **118**, 4527-39.

Carugo, O. and Argos, P. (1997). NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins*, **28**, 10-28.

Casal, E., Federici, L., Zhang, W., Fernandez-Recio, J., Priego, E. M., Miguel, R. N., DuHadaway, J. B., Prendergast, G. C., Luisi, B. F. and Laue, E. D. (2006). The crystal structure of the BAR domain from human Bin1/amphiphysin II and its implications for molecular recognition. *Biochemistry*, **45**, 12917-28.

Chen, F., Demers, L. M. and Shi, X. (2002). Upstream signal transduction of NF-kappaB activation. *Curr Drug Targets Inflamm Allergy*, **1**, 137-49.

Chistoserdova, L. V. and Lidstrom, M. E. (1991). Purification and characterization of hydroxypyruvate reductase from the facultative methylotroph Methylobacterium extorquens AM1. *J Bacteriol*, **173**, 7228-32.

Clarke, A. R. and Daffom, T. R. (1998). Nicotinamide cofactor-dependent enzymes. *Comprehensive Biology Catalysis: A Mechanistic Reference*. M. Sinnott. New York, Academic Press. **III:** 1-80.

Clarke, A. R., Wigley, D. B., Chia, W. N., Barstow, D., Atkinson, T. and Holbrook, J. J. (1986). Site-Directed Mutagenesis Reveals Role of Mobile Arginine Residue in Lactate-Dehydrogenase Catalysis. *Nature*, **324**, 699-702.

Collins, B. M., Norwood, S. J., Kerr, M. C., Mahony, D., Seaman, M. N., Teasdale, R. D. and Owen, D. J. (2008). Structure of Vps26B and mapping of its interaction with the retromer protein complex. *Traffic*, **9**, 366-79.

Collins, B. M., Skinner, C. F., Watson, P. J., Seaman, M. N. and Owen, D. J. (2005). Vps29 has a phosphoesterase fold that acts as a protein interaction scaffold for retromer assembly. *Nat Struct Mol Biol*, **12**, 594-602.

Conti, L. and Bradley, D. (2007). TERMINAL FLOWER1 is a mobile signal controlling Arabidopsis architecture. *Plant Cell*, **19**, 767-78.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science*, **316**, 1030-3.

Corbit, K. C., Trakul, N., Eves, E. M., Diaz, B., Marshall, M. and Rosner, M. R. (2003). Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein. *J Biol Chem*, **278**, 13061-8.

Coulson, C. J. and Rabin, B. R. (1969). Inhibition of lactate dehydrogenase by high concentrations of pyruvate: The nature and removal of the inhibitor. *FEBS Lett*, **3**, 333-337.

Coulter-Mackie, M. B. (2006). 4-Hydroxyproline metabolism and glyoxylate production: A target for substrate depletion in primary hyperoxaluria? *Kidney Int*, **70**, 1891-1893.

Cozier, G. E., Carlton, J., McGregor, A. H., Gleeson, P. A., Teasdale, R. D., Mellor, H. and Cullen, P. J. (2002). The phox homology (PX) domain-dependent, 3-phosphoinositide-mediated association of sorting nexin-1 with an early sorting endosomal compartment is required for its ability to regulate epidermal growth factor receptor degradation. *J Biol Chem*, **277**, 48730-6.

Cramer, S. D., Ferree, P. M., Lin, K., Milliner, D. S. and Holmes, R. P. (1999). The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II. *Hum Mol Genet*, **8**, 2063-9.

Cregeen, D. P., Williams, E. L., Hulton, S. and Rumsby, G. (2003). Molecular analysis of the glyoxylate reductase (GRHPR) gene and description of mutations underlying primary hyperoxaluria type 2. *Hum Mutat*, **22**, 497.

Cullen, P. J. (2008). Endosomal sorting and signalling: an emerging role for sorting nexins. *Nat Rev Mol Cell Biol*, **9**, 574-82.

D'Arcy, A., Villard, F. and Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr*, **63**, 550-4.

Danpure, C. J. (2005). Primary hyperoxaluria: from gene defects to designer drugs? *Nephrol Dial Transplant*, **20**, 1525-9.

Danpure, C. J. and Jennings, P. R. (1986). Peroxisomal alanine:glyoxylate aminotransferase deficiency in primary hyperoxaluria type I. *FEBS Lett*, **201**, 20-4.

Danpure, C. J. and Purdue, P. E. (1995). Primary hyperoxaluria. *The Metabolic and Molecular Basis of Inherited Disease*. C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle. New York, McGraw-Hill. **II:** 2385-2424.

Danpure, C. J. and Rumsby, G. (2004). Molecular aetiology of primary hyperoxaluria and its implications for clinical management. *Expert Rev Mol Med*, **6**, 1-16.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. and Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, **35**, W375-83.

Dawkins, P. D. and Dickens, F. (1965). Oxidation of D- and L-Glycerate by Rat Liver. *Biochem J*, **94**, 353-67.

DeLano, W. L. (2002). *The PyMOL molecular graphics system*. DeLano Scientific, San Carlos, CA, USA.

Dengler, U., Niefind, K., Kiess, M. and Schomburg, D. (1997). Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, NAD+ and 2-oxoisocaproate at 1.9 A resolution. *J Mol Biol*, **267**, 640-60.

Denicola-Seoane, A. and Anderson, B. M. (1990). Purification and characterization of Haemophilus influenzae D-lactate dehydrogenase. *J Biol Chem*, **265**, 3691-6.

Dirr, H., Reinemer, P. and Huber, R. (1994). X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur J Biochem*, **220**, 645-61.

Duncan, R. J. and Tipton, K. F. (1969). The oxidation and reduction of glyoxylate by lactic dehydrogenase. *Eur J Biochem*, **11**, 58-61.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-32.

Eszes, C. M., Sessions, R. B., Clarke, A. R., Moreton, K. M. and Holbrook, J. J. (1996). Removal of substrate inhibition in a lactate dehydrogenase from human muscle by a single residue change. *FEBS Lett*, **399**, 193-7.

Ferl, R. J. (2004). 14-3-3 proteins: regulation of signal-induced events. *Physiol Plant*, **120**, 173-178.

Fersht, A. (1985). Enzyme structure and mechanism. New York, Freeman.

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Banfield, M. J. and Rameau, C. (2003). DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell*, **15**, 2742-54.

Frederick, E. W., Smith, L. H., Rabkin, M. T. and Richie, R. H. (1963). Studies on Primary Hyperoxaluria .1. In Vivo Demonstration of a Defect in Glyoxylate Metabolism. *N Eng J Med*, **269**, 821-9.

Gallop, J. L. and McMahon, H. T. (2005). BAR domains and membrane curvature: bringing your curves to the BAR. *Biochem Soc Symp*, **72**, 223-31.

Garman, E. F. and Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **62**, 32-47.

Gibbs, D. A. and Watts, R. W. E. (1973). Identification of Enzymes That Catalyze Oxidation of Glyoxylate to Oxalate in 100000 G Supernatant Fraction of Human Hyperoxaluric and Control Liver and Heart Tissue. *Clin Sci*, **44**, 227-41.

Gillooly, D. J., Morrow, I. C., Lindsay, M., Gould, R., Bryant, N. J., Gaullier, J. M., Parton, R. G. and Stenmark, H. (2000). Localization of phosphatidylinositol 3-phosphate in yeast and mammalian cells. *Embo J*, **19**, 4577-88.

Givan, C. V. and Kleczkowski, L. A. (1992). The Enzymic Reduction of Glyoxylate and Hydroxypyruvate in Leaves of Higher Plants. *Plant Physiol*, **100**, 552-556.

Goldberg, J. D., Yoshida, T. and Brick, P. (1994). Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 A resolution. *J Mol Biol*, **236**, 1123-40.

Gomez, M. S., Piper, R. C., Hunsaker, L. A., Royer, R. E., Deck, L. M., Makler, M. T. and Vander Jagt, D. L. (1997). Substrate and cofactor specificity and selective inhibition of lactate dehydrogenase from the malarial parasite P. falciparum. *Mol Biochem Parasitol*, **90**, 235-46.

Grant, G. A. (1989). A new family of 2-hydroxyacid dehydrogenases. *Biochem Biophys Res Commun*, **165**, 1371-4.

Greenler, J. M., Sloan, J. S., Schwartz, B. W. and Becker, W. M. (1989). Isolation, Characterization and Sequence-Analysis of a Full-Length Cdna Clone Encoding Nadh-Dependent Hydroxypyruvate Reductase from Cucumber. *Plant Mol Biol*, **13**, 139-150.

Griffin, C. T., Trejo, J. and Magnuson, T. (2005). Genetic evidence for a mammalian retromer complex containing sorting nexins 1 and 2. *Proc Natl Acad Sci U S A*, **102**, 15173-7.

Gullapalli, A., Garrett, T. A., Paing, M. M., Griffin, C. T., Yang, Y. and Trejo, J. (2004). A role for sorting nexin 2 in epidermal growth factor receptor down-regulation: evidence for distinct functions of sorting nexin 1 and 2 in protein trafficking. *Mol Biol Cell*, **15**, 2143-55.

Gullapalli, A., Wolfe, B. L., Griffin, C. T., Magnuson, T. and Trejo, J. (2006). An essential role for SNX1 in lysosomal sorting of protease-activated receptor-1: evidence for retromer-, Hrs-, and Tsg101-independent functions of sorting nexins. *Mol Biol Cell*, **17**, 1228-38.

Habermann, B. (2004). The BAR-domain family of proteins: a case of bending and binding? *EMBO Rep*, **5**, 250-5.

Haft, C. R., de la Luz Sierra, M., Bafford, R., Lesniak, M. A., Barr, V. A. and Taylor, S. I. (2000). Human orthologs of yeast vacuolar protein sorting proteins Vps26, 29, and 35: assembly into multimeric complexes. *Mol Biol Cell*, **11**, 4105-16.

Hanzawa, Y., Money, T. and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proc Natl Acad Sci U S A*, **102**, 7748-53.

Hayward, S. and Berendsen, H. J. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*, **30**, 144-54.

Hayward, S. and Lee, R. A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model*, **21**, 181-3.

Heiseke, A., Schobel, S., Lichtenthaler, S. F., Vorberg, I., Groschup, M. H., Kretzschmar, H., Schatzl, H. M. and Nunziante, M. (2008). The novel sorting nexin SNX33 interferes with cellular PrP formation by modulation of PrP shedding. *Traffic*, **9**, 1116-29.

Henderson, I. R. and Dean, C. (2004). Control of Arabidopsis flowering: the chill before the bloom. *Development*, **131**, 3829-38.

Hewitt, C. O., Eszes, C. M., Sessions, R. B., Moreton, K. M., Dafforn, T. R., Takei, J., Dempsey, C. E., Clarke, A. R. and Holbrook, J. J. (1999). A general method for relieving substrate inhibition in lactate dehydrogenases. *Protein Eng*, **12**, 491-6.

Hierro, A., Rojas, A. L., Rojas, R., Murthy, N., Effantin, G., Kajava, A. V., Steven, A. C., Bonifacino, J. S. and Hurley, J. H. (2007). Functional architecture of the retromer cargo-recognition complex. *Nature*, **449**, 1063-7.

Holmes, R. P., Hurst, C. H., Assimos, D. G. and Goodman, H. O. (1995). Glucagon Increases Urinary Oxalate Excretion in the Guinea-Pig. *Am J Physiol*, **32**, 568-574.

Izumi, Y., Yoshida, T., Kanzaki, H., Toki, S., Miyazaki, S. S. and Yamada, H. (1990). Purification and characterization of hydroxypyruvate reductase from a serine-producing methylotroph, Hyphomicrobium methylovorum GM2. *Eur J Biochem*, **190**, 279-84.

Jaeger, K. E. and Wigge, P. A. (2007). FT protein acts as a long-range signal in Arabidopsis. *Curr Biol*, **17**, 1050-4.

Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. and Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends Plant Sci*, **7**, 106-11.

Jones, J. M., Morrell, J. C. and Gould, S. J. (2000). Identification and characterization of HAOX1, HAOX2, and HAOX3, three human peroxisomal 2-hydroxy acid oxidases. *J Biol Chem*, **275**, 12590-7.

Kamoda, N., Minatogawa, Y., Nakamura, M., Nakanishi, J., Okuno, E. and Kido, R. (1980). The Organ Distribution of Human Alanine-2-Oxoglutarate Aminotransferase and Alanine-Glyoxylate Aminotransferase. *Biochem Med*, **23**, 25-34.

Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., Chory, J., Harrison, M. J. and Weigel, D. (1999). Activation tagging of the floral inducer FT. *Science*, **286**, 1962-5.

Kleywegt, G. J. and Jones, T. A. (1994). A super position. *CCP4/ESF-EACBM newsletter on protein crystallography*, **31**, 9-14.

Knight, J., Holmes, R. P., Milliner, D. S., Monico, C. G. and Cramer, S. D. (2006a). Glyoxylate reductase activity in blood mononuclear cells and the diagnosis of primary hyperoxaluria type 2. *Nephrol Dial Transplant*, **21**, 2292-2295.

Knight, J., Jiang, J., Assimos, D. G. and Holmes, R. P. (2006b). Hydroxyproline ingestion and urinary oxalate and glycolate excretion. *Kidney Int*, **70**, 1929-1934.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. *Science*, **286**, 1960-2.

Kochhar, S., Chuard, N. and Hottinger, H. (1992). Glutamate 264 modulates the pH dependence of the NAD(+)-dependent D-lactate dehydrogenase. *J Biol Chem*, **267**, 20298-301.

Kochhar, S., Hunziker, P. E., Leong-Morgenthaler, P. and Hottinger, H. (1992a). Primary structure, physicochemical properties, and chemical modification of NAD(+)-

dependent D-lactate dehydrogenase. Evidence for the presence of Arg-235, His-303, Tyr-101, and Trp-19 at or near the active site. *J Biol Chem*, **267**, 8499-513.

Kochhar, S., Hunziker, P. E., Leongmorgenthaler, P. and Hottinger, H. (1992b). Evolutionary Relationship of Nad+-Dependent D-Lactate Dehydrogenase - Comparison of Primary Structure of 2-Hydroxy Acid Dehydrogenases. *Biochem Biophys Res Comm*, **184**, 60-66.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T. and Yano, M. (2002). Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol*, **43**, 1096-105.

Kumar, V., Carlson, J. E., Ohgi, K. A., Edwards, T. A., Rose, D. W., Escalante, C. R., Rosenfeld, M. G. and Aggarwal, A. K. (2002). Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase. *Mol Cell*, **10**, 857-69.

Kurten, R. C., Cadena, D. L. and Gill, G. N. (1996). Enhanced degradation of EGF receptors by a sorting nexin, SNX1. *Science*, **272**, 1008-10.

Kutzenko, A. S., Lamzin, V. S. and Popov, V. O. (1998). Conserved supersecondary structural motif in NAD-dependent dehydrogenases. *FEBS Lett*, **423**, 105-9.

Lamzin, V. S., Dauter, Z., Popov, V. O., Harutyunyan, E. H. and Wilson, K. S. (1994). High resolution structures of holo and apo formate dehydrogenase. *J Mol Biol*, **236**, 759-85.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.

Lee, J., Retamal, C., Cuitino, L., Caruano-Yzermans, A., Shin, J. E., van Kerkhof, P., Marzolo, M. P. and Bu, G. (2008). Adaptor protein sorting nexin 17 regulates amyloid precursor protein trafficking and processing in the early endosomes. *J Biol Chem*, **283**, 11501-8.

Lemmon, S. K. and Traub, L. M. (2000). Sorting in the endosomal system in yeast and animal cells. *Curr Opin Cell Biol*, **12**, 457-66.

Leumann, E. and Hoppe, B. (1999). What is new in primary hyperoxaluria? *Nephrol Dial Transplant*, **14**, 2556-2558.

Li, H. and Goldstein, B. M. (1992). Carboxamide group conformation in the nicotinamide and thiazole-4-carboxamide rings: implications for enzyme binding. *J Med Chem*, **35**, 3560-7.

Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S. and Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell*, **11**, 1007-18.

Liu, H., Liu, Z. Q., Chen, C. X., Magill, S., Jiang, Y. and Liu, Y. J. (2006). Inhibitory regulation of EGF receptor degradation by sorting nexin 5. *Biochem Biophys Res Commun*, **342**, 537-46.

Lluis, C. and Bozal, J. (1977). Relationship between hydroxypyruvate and the production of oxalate in vitro. *Biochim Biophys Acta*, **461**, 209-17.

Loo, T. W. and Clarke, D. M. (2007). Chemical and pharmacological chaperones as new therapeutic agents. *Expert Rev Mol Med*, **9**, 1-18.

Lorenz, K., Lohse, M. J. and Quitterer, U. (2003). Protein kinase C switches the Raf kinase inhibitor from Raf-1 to GRK-2. *Nature*, **426**, 574-9.

Ma, L. and Kovacs, J. A. (2000). Expression and characterization of recombinant human-derived Pneumocystis carinii dihydrofolate reductase. *Antimicrob Agents Chemother*, **44**, 3092-6.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F. S. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res*, **19**, 1154.

Martins, B. M., Macedo-Ribeiro, S., Bresser, J., Buckel, W. and Messerschmidt, A. (2005). Structural basis for stereo-specific catalysis in NAD(+)-dependent (R)-2-hydroxyglutarate dehydrogenase from *Acidaminococcus fermentans*. *FEBS J*, **272**, 269-81.

Masuda, M., Takeda, S., Sone, M., Ohki, T., Mori, H., Kamioka, Y. and Mochizuki, N. (2006). Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms. *Embo J*, **25**, 2889-97.

Mathews, C. K., van Holde, K. E. and Ahern, K. G. (2000). Biochemistry. San Francisco, Benjamin Cummings.

Mathieu, J., Warthmann, N., Kuttner, F. and Schmid, M. (2007). Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. *Curr Biol*, **17**, 1055-60.

McFerrin, M. B. and Snell, E. H. (2002). The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J Appl Crystallogr*, **35**, 538-545.

McMahon, H. T. and Gallop, J. L. (2005). Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature*, **438**, 590-6.

Mdluli, K., Booth, M. P., Brady, R. L. and Rumsby, G. (2005). A preliminary account of the properties of recombinant human Glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism. *Biochim Biophys Acta*, **1753**, 209-16.

Meany, J. E. and Pocker, Y. (1991). The dehydration of glyoxylate hydrate: General-acid, general-base, metal ion, and enzymatic catalysis. *J Am Chem Soc*, **113**, 6155-61.

Meister, A. (1952). Enzymatic Preparation of Alpha-Keto Acids. *J Biol Chem*, **197**, 309-317.

MGC project team (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res*, **14**, 2121-7.

Michel, G., Roszak, A. W., Sauve, V., Maclean, J., Matte, A., Coggins, J. R., Cygler, M. and Lapthorn, A. J. (2003). Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. *J Biol Chem*, **278**, 19463-72.

Miller, D. (2004). Structural Studies of Novel Sperm Related Proteins. Biochemistry Department. Bristol, University of Bristol.

Mistry, J., Danpure, C. J. and Chalmers, R. A. (1988). Hepatic D-Glycerate Dehydrogenase and Glyoxylate Reductase Deficiency in Primary Hyperoxaluria Type-2. *Biochem Soc Trans*, **16**, 626-627.

Monico, C. G., Persson, M., Ford, G. C., Rumsby, G. and Milliner, D. S. (2002). Potential mechanisms of marked hyperoxaluria not due to primary hyperoxaluria I or II. *Kidney Int*, **62**, 392-400.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-55.

Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, **247**, 536-40.

Nakagawa, M., Shimamoto, K. and Kyozuka, J. (2002). Overexpression of RCN1 and RCN2, rice TERMINAL FLOWER 1/CENTRORADIALIS homologs, confers delay of phase transition and altered panicle morphology in rice. *Plant J*, **29**, 743-50.

Nardini, M., Spano, S., Cericola, C., Pesce, A., Massaro, A., Millo, E., Luini, A., Corda, D. and Bolognesi, M. (2003). CtBP/BARS: a dual-function protein involved in transcription co-repression and Golgi membrane fission. *Embo J*, **22**, 3122-30.

Neilands, J. B. (1954). Studies on lactic dehydrogenase of heart. III. Action of inhibitors. *J Biol Chem*, **208**, 225-30.

Newsholme, E. A. and Leech, A. R. (1983). Biochemistry for the Medical Sciences, John Wiley and Sons Ltd.

Nisselbaum, J. S. and Bodansky, O. (1963). Purification, Kinetic, and Immunochemical Studies of Major Variants of Lactic Dehydrogenase from Human Liver, Hepatoma, and

Erythrocytes - Comparison with Major Variant of Human Heart Lactic Dehydrogenase. *J Biol Chem*, **238**, 969-74.

Oganesyan, N., Kim, S. H. and Kim, R. (2005). On-column protein refolding for crystallization. *J Struct Funct Genomics*, **6**, 177-82.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode, Academic Press.

Painter, J. and Merritt, M. A. (2006). TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr*, **39**, 109-111.

Perrakis, A., Morris, R. and Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol*, **6**, 458-463.

Peter, B. J., Kent, H. M., Mills, I. G., Vallis, Y., Butler, P. J., Evans, P. R. and McMahon, H. T. (2004). BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. *Science*, **303**, 495-9.

Pikielny, C. W., Hasan, G., Rouyer, F. and Rosbash, M. (1994). Members of a family of Drosophila putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron*, **12**, 35-49.

Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N. and Lifschitz, E. (2001). Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. *Plant Cell*, **13**, 2687-702.

Poore, R. E., Hurst, C. H., Assimos, D. G. and Holmes, R. P. (1997). Pathways of hepatic oxalate synthesis and their regulation. *Am J Physiol Cell Physiol*, **41**, C289-C294.

Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. and Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr*, **58**, 1955-7.

Pylypenko, O., Ignatev, A., Lundmark, R., Rasmuson, E., Carlsson, S. R. and Rak, A. (2008). A combinatorial approach to crystallization of PX-BAR unit of the human Sorting Nexin 9. *J Struct Biol*, **162**, 356-60.

Pylypenko, O., Lundmark, R., Rasmuson, E., Carlsson, S. R. and Rak, A. (2007). The PX-BAR membrane-remodeling unit of sorting nexin 9. *Embo J*, **26**, 4788-800.

Raghavan, K. G., Lathika, K. M., Gandhi, N. M., D'Souza, S. J., Tarachand, U., Ramakrishnan, V. and Singh, B. B. (1997). Biogenesis of L-glyceric aciduria, oxalosis and renal injury in rats simulating type II primary hyperoxaluria. *Biochim Biophys Acta*, **1362**, 97-102.

Raghavan, K. G. and Richardson, K. E. (1983a). Hydroxypyruvate-Mediated Regulation of Oxalate Synthesis by Lactate-Dehydrogenase and Its Relevance to Primary Hyperoxaluria Type-Ii. *Biochem Med*, **29**, 101-113.

Raghavan, K. G. and Richardson, K. E. (1983b). Hyperoxaluria in L-Glyceric Aciduria - Possible Non-Enzymic Mechanism. *Biochem Med*, **29**, 114-121.

Rao, S. T. and Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *J Mol Biol*, **76**, 241-56.

Rath, O., Park, S., Tang, H. H., Banfield, M. J., Brady, R. L., Lee, Y. C., Dignam, J. D., Sedivy, J. M., Kolch, W. and Yeung, K. C. (2008). The RKIP (Raf-1 Kinase Inhibitor Protein) conserved pocket binds to the phosphorylated N-region of Raf-1 and inhibits the Raf-1-mediated activated phosphorylation of MEK. *Cell Signal*, **20**, 935-41.

Razeto, A., Kochhar, S., Hottinger, H., Dauter, M., Wilson, K. S. and Lamzin, V. S. (2002). Domain closure, substrate specificity and catalysis of D-lactate dehydrogenase from Lactobacillus bulgaricus. *J Mol Biol*, **318**, 109-19.

Read, J. A., Winter, V. J., Eszes, C. M., Sessions, R. B. and Brady, R. L. (2001). Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins*, **43**, 175-85.

Rendina, A. R., Hermes, J. D. and Cleland, W. W. (1984). A Novel Method for Determining Rate Constants for Dehydration of Aldehyde Hydrates. *Biochemistry*, **23**, 5148-5156.

Richardson, K. E. and Tolbert, N. E. (1961). Oxidation of glyoxylic acid to oxalic acid by glycolic acid oxidase. *J Biol Chem*, **236**, 1280-4.

Roberts, M. R. (2003). 14-3-3 proteins find new partners in plant cell signalling. *Trends Plant Sci*, **8**, 218-23.

Romano, M. and Cerra, M. (1969). The action of crystalline lactate dehydrogenase from rabbit muscle on glyoxylate. *Biochim Biophys Acta*, **177**, 421-6.

Rumsby, G. (2006). Is liver analysis still required for the diagnosis of primary hyperoxaluria type 2? *Nephrol Dial Transplant*, **21**, 2063-2064.

Rumsby, G. and Cregeen, D. P. (1999). Identification and expression of a cDNA for human hydroxypyruvate/glyoxylate reductase. *Biochim Biophys Acta*, **1446**, 383-8.

Rumsby, G., Sharma, A., Cregeen, D. P. and Solomon, L. R. (2001). Primary hyperoxaluria type 2 without L-glycericaciduria: is the disease under-diagnosed? *Nephrol Dial Transplant*, **16**, 1697-9.

Rumsby, G., Williams, E. and Coulter-Mackie, M. (2004). Evaluation of mutation screening as a first line test for the diagnosis of the primary hyperoxalurias. *Kidney Int*, **66**, 959-963.

Samach, A., Onouchi, H., Gold, S. E., Ditta, G. S., Schwarz-Sommer, Z., Yanofsky, M. F. and Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science*, **288**, 1613-6.

Sawaki, S., Hattori, H. and Yamada, K. (1967). Glyoxylate Dehydrogenase Activity of Lactate Dehydrogenase. *J Biochem*, **62**, 263-8.

Sawaki, S. and Yamada, K. (1966). Glyoxylate reductase activity of lactate dehydrogenase. *Nature*, **210**, 91.

Schmitt, J., Hess, H. and Stunnenberg, H. G. (1993). Affinity purification of histidine-tagged proteins. *Mol Biol Rep*, **18**, 223-30.

Schuller, D. J., Grant, G. A. and Banasak, L. J. (1995). The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. *Nat Struct Biol*, **2**, 69-76.

Seaman, M. N., Marcusson, E. G., Cereghino, J. L. and Emr, S. D. (1997). Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J Cell Biol*, **137**, 79-92.

Seaman, M. N., McCaffery, J. M. and Emr, S. D. (1998). A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. *J Cell Biol*, **142**, 665-81.

Seaman, M. N. and Williams, H. P. (2002). Identification of the functional domains of yeast sorting nexins Vps5p and Vps17p. *Mol Biol Cell*, **13**, 2826-40.

Seet, L. F. and Hong, W. (2006). The Phox (PX) domain proteins and membrane traffic. *Biochim Biophys Acta*, **1761**, 878-96.

Serre, L., Pereira de Jesus, K., Zelwer, C., Bureaud, N., Schoentgen, F. and Benedetti, H. (2001). Crystal structures of YBHB and YBCL from Escherichia coli, two bacterial homologues to a Raf kinase inhibitor protein. *J Mol Biol*, **310**, 617-34.

Serre, L., Vallee, B., Bureaud, N., Schoentgen, F. and Zelwer, C. (1998). Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. *Structure*, **6**, 1255-65.

Shi, H., Rojas, R., Bonifacino, J. S. and Hurley, J. H. (2006). The retromer subunit Vps26 has an arrestin fold and binds Vps35 through its C-terminal domain. *Nat Struct Mol Biol*, **13**, 540-548.

Shimada, A., Niwa, H., Tsujita, K., Suetsugu, S., Nitta, K., Hanawa-Suetsugu, K., Akasaka, R., Nishino, Y., Toyama, M., Chen, L., Liu, Z. J., Wang, B. C., Yamamoto, M., Terada, T., Miyazawa, A., Tanaka, A., Sugano, S., Shirouzu, M., Nagayama, K., Takenawa, T. and Yokoyama, S. (2007). Curved EFC/F-BAR-domain dimers are joined end to end into a filament for membrane invagination in endocytosis. *Cell*, **129**, 761-72.

Shinoda, T., Arai, K. and Taguchi, H. (2007). A highly specific glyoxylate reductase derived from a formate dehydrogenase. *Biochem Biophys Res Comm*, **355**, 782-787.

Shoemark, D. K., Cliff, M. J., Sessions, R. B. and Clarke, A. R. (2007). Enzymatic properties of the lactate dehydrogenase enzyme from Plasmodium falciparum. *Febs J*, **274**, 2738-48.

Simister, P. C. (2004). Structural Studies of Raf Kinase Inhibitor Protein (RKIP) Family. Biochemistry Department. Bristol, University of Bristol.

Simister, P. C., Banfield, M. J. and Brady, R. L. (2002). The crystal structure of PEBP-2, a homologue of the PEBP/RKIP family. *Acta Crystallogr D Biol Crystallogr*, **58**, 1077-80.

Song, J., Zhao, K. Q., Newman, C. L., Vinarov, D. A. and Markley, J. L. (2007). Solution structure of human sorting nexin 22. *Protein Sci*, **16**, 807-14.

Soulet, F., Yarar, D., Leonard, M. and Schmid, S. L. (2005). SNX9 regulates dynamin assembly and is required for efficient clathrin-mediated endocytosis. *Mol Biol Cell*, **16**, 2058-67.

Stoll, V. S., Kimber, M. S. and Pai, E. F. (1996). Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase. *Structure*, **4**, 437-447.

Storoni, L. C., McCoy, A. J. and Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr*, **60**, 432-8.

Taguchi, H. and Ohta, T. (1993). Histidine-296 Is Essential for the Catalysis in Lactobacillus-Plantarum D-Lactate Dehydrogenase. *J Biol Chem*, **268**, 18030-18034.

Taguchi, H. and Ohta, T. (1994). Essential Role of Arginine-235 in the Substrate-Binding of Lactobacillus-Plantarum D-Lactate Dehydrogenase. *J Biochem*, **115**, 930-936.

Taguchi, H., Ohta, T. and Matsuzawa, H. (1997). Involvement of Glu-264 and Arg-235 in the essential interaction between the catalytic imidazole and substrate for the D-lactate dehydrogenase catalysis. *J Biochem*, **122**, 802-809.

Takada, S. and Goto, K. (2003). Terminal flower2, an Arabidopsis homolog of heterochromatin protein1, counteracts the activation of flowering locus T by constans in the vascular tissues of leaves to regulate flowering time. *Plant Cell*, **15**, 2856-65.

Takayama, T., Nagata, M., Ozono, S., Nonomura, K. and Cramer, S. D. (2007). A novel mutation in the GRHPR gene in a Japanese patient with primary hyperoxaluria type 2. *Nephrol Dial Transplant*, **22**, 2371-2374.

Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S. and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science*, **316**, 1033-6.

Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A. and Mitsui, Y. (1996). Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 A

resolution: the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. *Structure*, **4**, 33-45.

Tarricone, C., Xiao, B., Justin, N., Walker, P. A., Rittinger, K., Gamblin, S. J. and Smerdon, S. J. (2001). The structural basis of Arfaptin-mediated cross-talk between Rac and Arf signalling pathways. *Nature*, **411**, 215-9.

Teasdale, R. D., Loci, D., Houghton, F., Karlsson, L. and Gleeson, P. A. (2001). A large family of endosome-localized proteins related to sorting nexin 1. *Biochem J*, **358**, 7-16.

Tohdoh, N., Tojo, S., Agui, H. and Ojika, K. (1995). Sequence homology of rat and human HCNP precursor proteins, bovine phosphatidylethanolamine-binding protein and rat 23-kDa protein associated with the opioid-binding protein. *Brain Res Mol Brain Res*, **30**, 381-4.

Vaguine, A. A., Richelle, J. and Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr*, **55**, 191-205.

van Schaftingen, E., Draye, J. P. and van Hoof, F. (1989). Coenzyme specificity of mammalian liver D-glycerate dehydrogenase. *Eur J Biochem*, **186**, 355-9.

Verleur, N. and Wanders, R. J. A. (1993). Permeability Properties of Peroxisomes in Digitonin-Permeabilized Rat Hepatocytes - Evidence for Free Permeability Towards a Variety of Substrates. *Eur J Biochem* **218**, 75-82.

Vieille, C. and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*, **65**, 1-43.

Vinals, C., De Bolle, X., Depiereux, E. and Feytmans, E. (1995). Knowledge-based modeling of the D-lactate dehydrogenase three-dimensional structure. *Proteins*, **21**, 307-18.

Vinals, C., Depiereux, E. and Feytmans, E. (1993). Prediction of structurally conserved regions of D-specific hydroxy acid dehydrogenases by multiple alignment with formate dehydrogenase. *Biochem Biophys Res Commun*, **192**, 182-8.

Wang, D., Guo, M., Liang, Z., Fan, J., Zhu, Z., Zang, J., Li, X., Teng, M., Niu, L., Dong, Y. and Liu, P. (2005). Crystal structure of human vacuolar protein sorting protein 29 reveals a phosphodiesterase/nuclease-like fold and two protein-protein interaction sites. *J Biol Chem*, **280**, 22962-7.

Warren, W. A. (1970). Catalysis of Both Oxidation and Reduction of Glyoxylate by Pig Heart Lactate Dehydrogenase Isozyme-1. *J Biol Chem*, **245**, 1675-&.

Wassmer, T., Attar, N., Bujny, M. V., Oakley, J., Traer, C. J. and Cullen, P. J. (2007). A loss-of-function screen reveals SNX5 and SNX6 as potential components of the mammalian retromer. *J Cell Sci*, **120**, 45-54.

Watts, R. W. E., Chalmers, R. A., Gibbs, D. A., Lawson, A. M., Purkiss, P. and Spellacy, E. (1979). Studies on Some Possible Biochemical Treatments of Primary Hyperoxaluria. *Q J Med*, **48**, 259-272.

Webster, K. E., Ferree, P. M., Holmes, R. P. and Cramer, S. D. (2000). Identification of missense, nonsense, and deletion mutations in the GRHPR gene in patients with primary hyperoxaluria type II (PH2). *Hum Genet*, **107**, 176-85.

Weissenhorn, W. (2005). Crystal structure of the endophilin-A1 BAR domain. *J Mol Biol*, **351**, 653-61.

Wierenga, R. K., Demaeyer, M. C. H. and Hol, W. G. J. (1985). Interaction of Pyrophosphate Moieties with Alpha-Helixes in Dinucleotide Binding-Proteins. *Biochemistry*, **24**, 1346-1357.

Wierenga, R. K. and Hol, W. G. J. (1983). Predicted Nucleotide-Binding Properties of P21 Protein and Its Cancer-Associated Variant. *Nature*, **302**, 842-844.

Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U. and Weigel, D. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. *Science*, **309**, 1056-9.

Wigley, D. B., Gamblin, S. J., Turkenburg, J. P., Dodson, E. J., Piontek, K., Muirhead, H. and Holbrook, J. J. (1992). Structure of a ternary complex of an allosteric lactate dehydrogenase from Bacillus stearothermophilus at 2.5 A resolution. *J Mol Biol*, **223**, 317-35.

Williams, H. E., Johnson, G. A. and Smith, L. H. (1971). Renal Clearance of Oxalate in Normal Subjects and Patients with Primary Hyperoxaluria. *Clin Sci*, **41**, 213-8.

Williams, H. E. and Smith, L. H. (1968a). Identification and Determination of Glyceric Acid in Human Urine. *J Lab Clin Med*, **71**, 495-500.

Williams, H. E. and Smith, L. H. (1968b). L-Glyceric Aciduria - a New Genetic Variant of Primary Hyperoxaluria. *N Eng J Med*, **278**, 233-8.

Williams, H. E. and Smith, L. H. (1971). Hyperoxaluria in L-Glyceric Aciduria - Possible Pathogenic Mechanism. *Science*, **171**, 390-1.

Worby, C. A. and Dixon, J. E. (2002). Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol*, **3**, 919-31.

Xue, H. H., Sakaguchi, T., Fujie, M., Ogawa, H. and Ichiyama, A. (1999). Flux of the L-serine metabolism in rabbit, human, and dog livers - Substantial contributions of both mitochondrial and peroxisomal serine : pyruvate/alanine : glyoxylate aminotransferase. *J Biol Chem*, **274**, 16028-16033.

Yanagawa, M., Maedanakai, E., Yamakawa, K., Yamamoto, I., Kawamura, J., Tada, S. and Ichiyama, A. (1990). The Formation of Oxalate from Glycolate in Rat and Human Liver. *Biochimica Et Biophysica Acta*, **1036**, 24-33.

Yeung, K., Janosch, P., McFerran, B., Rose, D. W., Mischak, H., Sedivy, J. M. and Kolch, W. (2000). Mechanism of suppression of the Raf/MEK/extracellular signal-regulated kinase pathway by the raf kinase inhibitor protein. *Mol Cell Biol*, **20**, 3079-85.

Yeung, K., Seitz, T., Li, S., Janosch, P., McFerran, B., Kaiser, C., Fee, F., Katsanakis, K. D., Rose, D. W., Mischak, H., Sedivy, J. M. and Kolch, W. (1999). Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature*, **401**, 173-7.

Yoshikawa, S., Arai, R., Kinoshita, Y., Uchikubo-Kamo, T., Wakamatsu, T., Akasaka, R., Masui, R., Terada, T., Kuramitsu, S., Shirouzu, M. and Yokoyama, S. (2007). Structure of archaeal glyoxylate reductase from Pyrococcus horikoshii OT3 complexed with nicotinamide adenine dinucleotide phosphate. *Acta Crystallogr D Biol Crystallogr*, **63**, 357-65.

Zhang, X. X., Roe, S. M., Hou, Y. W., Bartlam, M., Rao, Z. H., Pearl, L. H. and Danpure, C. J. (2003). Crystal structure of alanine: Glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. *J Mol Biol*, **331**, 643-652.

Zhong, Q., Lazar, C. S., Tronchere, H., Sato, T., Meerloo, T., Yeo, M., Songyang, Z., Emr, S. D. and Gill, G. N. (2002). Endosomal localization and function of sorting nexin 1. *Proc Natl Acad Sci U S A*, **99**, 6767-72.

Zhong, Q., Watson, M. J., Lazar, C. S., Hounslow, A. M., Waltho, J. P. and Gill, G. N. (2005). Determinants of the endosomal localization of sorting nexin 1. *Mol Biol Cell*, **16**, 2049-57.

Zhou, C. Z., de La Sierra-Gallay, I. L., Quevillon-Cheruel, S., Collinet, B., Minard, P., Blondeau, K., Henckes, G., Aufrere, R., Leulliot, N., Graille, M., Sorel, I., Savarin, P., de la Torre, F., Poupon, A., Janin, J. and van Tilbeurgh, H. (2003). Crystal structure of the yeast Phox homology (PX) domain protein Grd19p complexed to phosphatidylinositol-3-phosphate. *J Biol Chem*, **278**, 50371-6.